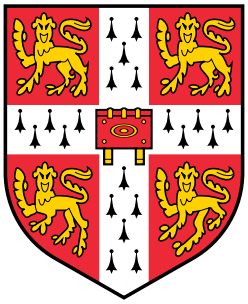
Large-scale density functional theory study of van-der-Waals heterostructures



Gabriel Constantinescu

Wolfson College

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

March 2018

Contents

List of Figures

List of Tables

Abstract

Research on two-dimensional (2D) materials currently occupies a sizeable fraction of the materials science community, which has led to the development of a comprehensive body of knowledge on such layered structures. However, the goal of this thesis is to deepen the understanding of the comparatively unknown heterostructures composed of different stacked layers. First, we utilise linear-scaling density functional theory (LS-DFT) to simulate intricate interfaces between the most promising layered materials, such as transition metal dichalcogenides (TMDC) or black phosphorus (BP) and hexagonal boron nitride (hBN). We show that hBN can protect BP from external influences, while also preventing the band-gap reduction in BP stacks, and enabling the use of BP heterostructures as tunnelling field effect transistors. Moreover, our simulations of the electronic structure of TMDC interfaces have reproduced photoemission spectroscopy observations, and have also provided an explanation for the coexistence of commensurate and incommensurate phases within the same crystal. Secondly, we have developed new functionality to be used in the future study of 2D heterostructures, in the form of a linear-response phonon formalism for LS-DFT. As part of its implementation, we have solved multiple implementation and theoretical issues through the use of novel algorithms.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the prescribed 60,000 word limit of the Degree Committee for the Faculty of Physics and Chemistry.

Acknowledgements

I would like to express my most sincere gratitude and admiration for Dr Nicholas Hine, who has provided unlimited guidance and mentorship throughout my doctoral degree. Moreover, I am thankful for the invaluable resources provided by Professor Mike Payne.

I am most grateful to my loving parents and family for their continuous support and for inspiring me to pursue my dreams.

Lastly, I truly appreciate the scholarships offered by the "Winton Programme for the Physics of Sustainability" and the Cambridge Trust, with the help of which I was able to perform research without the burden of financial concerns.

Chapter 1

Introduction

My doctoral research concerns the computational study of layered material interfaces through the use of linear-scaling density functional theory (LS-DFT). In turn, this branches out into simulations of electronic properties of 2D-material combinations, and the development of LS-DFT functionality. My thesis also reflects this segregation, with the first part being dedicated to the study of two-dimensional heterostructures, containing hexagonal boron nitride (hBN), black phosphorus (BP) and transition metal dichalcogenides (TMDCs), while the second part concerns the implementation of linear-response phonons in ONETEP [1].

In order to put my research into context, a brief account of the research on 2D materials is necessary. The first experimentally-obtained atomically thin layers consisted of graphene[2], which instantly caught the attention of the scientific community, due to its enormous electron mobility[3] ($1,000,000$ cm^2/Vs) and high mechanical resistance[4]. However, its applicability was severely limited by the lack of a band gap. In turn, this determined researchers to turn their attention to semiconducting inorganic layered materials, the most notable class being transition metal dichalcogenides [5](TMDCs), such as MoS_2, $MoSe_2$, WS_2, and WSe_2. TMDCs soon proved their worth in domains such as semiconductor electronics [6], spintronics[7], and optoelectronics[8]. This applicability range of TMDCs was ensured by their large current on/off ratios (due to large band gaps such as 1.9 eV for monolayer MoS_2), the presence of strong spin-orbit coupling in stacks without mirror symmetry [7], mobilities (400 cm^2/Vs for MoS_2 [9]) of the order of those encountered in silicon, and the ability to withstand large strains [10]. However, even TMDCs have flaws: most have a direct band gap only in monolayer form [11], the carrier mobilities are far smaller than those encountered in graphene, and their large band gap implies that the infrared (IR) spectrum cannot be harnessed by TMDC optoelectronics.

Consequently, the spotlight has recently been shared with layered black phosphorus (BP)[12, 13], a material which rectifies the most important shortcomings of TMDCs. The band gap of BP is direct even in stacked forms, and its value changes significantly with the number of layers: from 0.3 eV in bulk to 1.5-2.0 eV for the monolayer [14, 15]. This range of values allows BP to exploit a range of spectrum previously unharnessed by graphene and TMDCs, corresponding to the mid- and near-IR. Moreover, the higher carrier mobilities ($1,000$ cm^2/Vs)[16] compared to TMDC monolayers, the strongly anisotropic electronic and optical properties [17, 18, 19, 20, 21], as well as the robustness under elastic strain [22, 23, 24] have already ensured novel uses of BP in electronic [16, 18, 15, 25, 26], photonic [27, 28, 29], and thermoelectric [30, 31] applications.

Unfortunately, alongside these highly desirable properties, BP also has drawbacks. Firstly, it is highly sensitive to molecules present in air [32] and it is structurally unstable in ambient conditions [33, 34], meaning that pristine BP flakes are unusable after only a couple of hours [35]. Such effects prevent the holes in monolayer BP from reaching their maximum predicted mobility of $10,000$ cm^2/Vs [17]. Secondly, the decrease of the band gap in BP stacks implies that only the monolayer is usable for near-IR optoelectronic applications, limiting the external quantum efficiency of such devices. This band gap decrease cannot be avoided by simply misaligning the BP layers, as we have shown in a previous article [36] and in my Master dissertation [37] that stacked layers always interact at the Γ point of the Brilloin zone, exactly where the BP direct band-gap is. Lastly, due to the gap being smaller than that of TMDCs, the current on/off ratio for BP transistors is lower, making pristine BP channels not as appealing as TMDCs in ultra-low power applications.

Therefore, in the first part of Chapter 3 we propose that monolayers of an inert insulator, such as hexagonal boron nitride (hBN), could alleviate the aforementioned shortcomings of pristine BP. By simulating large rotated hBN/BP interfaces, we arrive at three conclusions: hBN encapsulation does not negatively affect the main electronic properties of BP, hBN spacers prevent the band gap reduction in stacked forms of BP, and BP/hBN/BP heterostructures are potential tunnelling field-effect transistors (TFETs), useful in both high-speed and low-power applications. The first point suggests that there is no electronic-structure downside to hBN-encapsulation, which is already known to pro-

tect BP from environmental interactions [38, 35, 39]. The second conclusion is especially relevant for optoelectronics, implying that hBN-spaced BP stacks have similar light-absorption properties as monolayer BP, with the added advantage of more photons being absorbed due to the increased thickness. Finally, since the operation of BP/hBN/BP as a TFET is based on quantum tunnelling, instead of thermal excitation of carriers, we encounter negative differential resistance (NDR) peaks [40]; their peak-to-valley ratios (PVRs) are comparable to those predicted for TMDC TFETs [41], and their subthreshold swings are below the minimum theoretical limit for conventional field effect transistors [42]. We note that intricate heterostructure 2D stacks have already been experimentally obtained [43], meaning that our proposition of hBN-encapsulated or spaced BP is realistic. However, our hBN/BP heterostructures are too large for traditional DFT to be an option. Therefore, for this study we utilise LS-DFT, as implemented in the ONETEP code [1], the framework of which we briefly describe in Chapter 2.

Chapter 3 concludes with the account of our study on bandstructure effects that occur when stacking TMDC layers, which was undertaken in conjunction with angular resolved photoemission spectroscopy (ARPES) performed by experimental collaborators [44]. Firstly, we have proven that LS-DFT can correctly account for hybridisation and band offsets in TMDC heterojunctions in good agreement with experiment. Moreover, we have been able to explain the presence of nonintuitive additional bands present in ARPES results, indicative of the coexistence of commensurate and incommensurate domains in slightly twisted $MoSe_2/WSe_2$ interfaces.

In the second part of the thesis, namely Chapter 4, we present the implementation of new functionality in ONETEP. One of the long term goals of the ONETEP community is to enable the computation of Raman spectra in a linear-scaling fashion. In the past 10 years, Raman spectroscopy has become a crucial multi-purpose investigation tool for layered materials [45]. In particular, relating to the previously discussed black phosphorus, Raman spectroscopy can determine the crystal orientation in micro-sized flakes, as the Raman active modes have different polarisation dependencies, strongly related to the crystal orientation [46]. Raman can also be used to determine the number of BP layers [47], as the interlayer breathing mode is highly redshifted with increasing thickness. Such

methods retain their usefulness throughout a wide range of 2D materials, such as TMDCs and graphene, as one can predict the number of layers or the stacking arrangements in a flake [45, 48, 49].

As the readers will see in this thesis, and from previous work [36], 2D heterostructures can form large coincidence cells, due to differing lattice parameters. Therefore, even though many traditional DFT codes have access to Raman spectroscopy simulations, such an approach would be too expensive for simulating large 2D heterostructures. Therefore, the main goal of the implementation part of my PhD work will be to create the framework for future linear-scaling Raman spectroscopy, through the implementation of accurate phonon spectra. Therefore, in Chapter 4 of this thesis I present my novel linear-scaling implementation of linear-response phonons in ONETEP, within the projector augmented wave (PAW [50]) formalism, for insulators.

While the well-known density-functional perturbation theory framework is followed for part of the derivations, the ONETEP constraints of linear-scaling and localised orbitals impose the need for an alternative approaches. Therefore, the novelty of our work is in developing an entirely new optimisation routine for the first-order wavefunctions, which consists of a combination of a preconditioned quasi-Newton approach (for the response kernel) and a preconditioned conjugate-gradient one (for the localised orbitals). Moreover, localisation constraints also resulted in the need to use a different displacement basis, meaning that for each perturbation we simultaneously move all atoms in certain directions, with certain associated weights. This framework also allows a trivial extension to electric field perturbations, which brings us closer to the ultimate goal of enabling fast and accurate Raman spectroscopy simulations in the context of 2D heterostructures.

Finally, in Chapter 5 I conclude my work, and briefly present ideas for future research.

Chapter 2

Theoretical considerations

In this chapter, we briefly discuss the general theory behind the ONETEP[1] implementation of linear-scaling Density Functional Theory (LS-DFT), and the projector-augmented wave (PAW) formalism [50]. The reader is assumed to be familiar with the central features of Density Functional Theory [51, 52], a full coverage of which can be found in Ref. [53].

2.1 Linear-Scaling DFT (LS-DFT) : ONETEP

2.1.1 Limitations of KS DFT

In traditional DFT one uses the Kohn-Sham formalism, in which the fully interacting system is mapped onto a non-interacting system with the same charge density. In this approach, the energy functional is minimised under the constraint that the Kohn-Sham (KS) orbitals, describing the electron states, are orthonormal. Thus, assuming that the number of KS orbitals is N, the number of orbital pairs that are required to be orthonormal scales as $O(N^2)$.

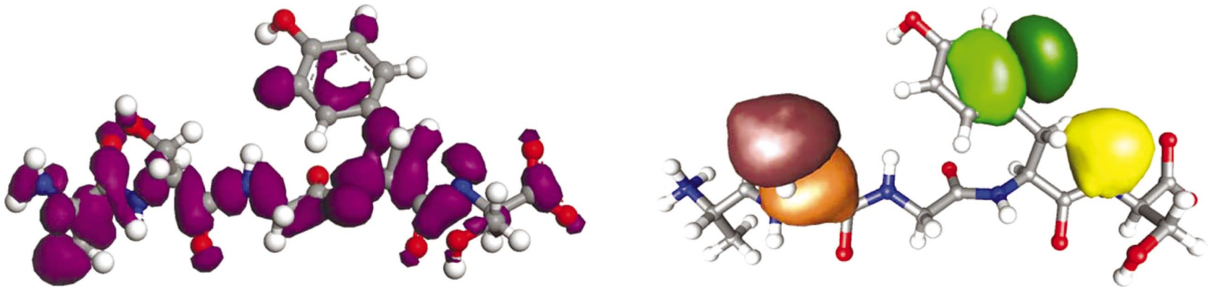


Figure 2.1: Left, one delocalized orbital $\psi_i(r)$ from a conventional DFT calculation [54] on a peptide. Right, three NGWFs from a ONETEP calculation on the same peptide [1].

As seen in Fig. 2.1, such KS orbitals also usually extend over the entire system, meaning that the extent of each orbital scales with the system size as $O(N)$. Combining the

previous scaling relations, a traditional KS DFT calculation scales as $O(N^3)$. While this does not pose a problem for relatively small systems, one encounters a scaling wall after a few hundred atoms. Since we will be investigating heterostructures containing more than 1500 atoms (Fig. 3.2.f,g), traditional DFT is simply not an option, and we require a different approach.

2.1.2 Density-matrix methods

One solution originates from the first order density matrix, which can fully represent the system properties according to the Hohenberg-Kohn theorems [51]:

$$\rho(\mathbf{r},\mathbf{r}') = \sum_n f_n \psi_n^*(\mathbf{r})\psi_n(\mathbf{r}') \ , \tag{2.1.1}$$

where f_n is the occupation number (either 0 or 1); note that and $\rho(\mathbf{r},\mathbf{r}) = \rho(\mathbf{r})$ is simply the charge density. However, the aforementioned problematic KS-orbital constraints are now translated into the idempotency condition for the density matrix:

$$\begin{aligned}\rho^2(\mathbf{r},\mathbf{r}') = &\sum_{i,j} f_i f_j \langle\mathbf{r}|\psi_i\rangle \langle\psi_i| \, |\psi_j\rangle \langle\psi_j|\mathbf{r}'\rangle = \sum_{i,j} f_i f_j \langle\mathbf{r}|\psi_i\rangle \, \delta_{ij} \, \langle\psi_j|\mathbf{r}'\rangle = \\ &\sum_k f_k \langle\mathbf{r}|\psi_k\rangle \langle\psi_k|\mathbf{r}'\rangle = \rho(\mathbf{r},\mathbf{r}') \ ,\end{aligned} \tag{2.1.2}$$

where we have implicitly used the orthonormality of the orbitals $|\psi_i\rangle$ and the fact that $f_i f_j$ is still either 0 or 1.

However, the information contained in the density matrix still scales as $O(N^2)$, due to orbital delocalisation and linear scaling with system size. The solution comes in the form of the near-sightedness principle [55, 56, 57], which states that in insulating or semiconducting systems the terms in the density matrix decay exponentially:

$$\rho(\mathbf{r},\mathbf{r}') \propto e^{-\gamma|\mathbf{r}-\mathbf{r}'|} \ , \tag{2.1.3}$$

implying that the density matrix is sparse. This also allows one to use localised orbitals accurately; thus, in ONETEP the density matrix is written as:

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{\alpha,\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r}') \ , \tag{2.1.4}$$

where $\phi_\alpha(\mathbf{r})$ is a localised orbital (Fig. 2.1), henceforth called non-orthogonal generalized Wannier function (NGWF) [1, 58], and $K^{\alpha\beta}$ is a density kernel [59, 1]. Thus, the square of the density matrix (Eq. 2.1.2) is now:

$$\begin{aligned}\rho^2(\mathbf{r},\mathbf{r}') &= \sum_{\alpha,\delta} \langle \mathbf{r}|\phi_\alpha\rangle \, K^{\alpha\delta} \, \langle \phi_\delta| \left[\sum_{\gamma,\beta} |\phi_\gamma\rangle \, K^{\gamma\beta} \, \langle \phi_\beta|\mathbf{r}'\rangle \right] = \\ &= \sum_{\alpha,\beta} \langle \mathbf{r}|\phi_\alpha\rangle \left[\sum_{\delta,\gamma} K^{\alpha\delta} S_{\delta\gamma} K^{\gamma\beta} \right] \langle \phi_\beta|\mathbf{r}'\rangle \ ,\end{aligned} \tag{2.1.5}$$

where $S_{\delta\gamma} = \langle \phi_\delta|\phi_\gamma\rangle$ is the overlap term between NGWFs ϕ_δ and ϕ_γ. If one requires that Eq. 2.1.2 is fulfilled, the square bracket of the previous equation needs to be equal to $K^{\alpha\beta}$:

$$\mathbf{KSK} = \mathbf{K} \tag{2.1.6}$$

The contravariant nature of the kernel $K^{\alpha\beta}$ makes it strictly related to the NGWF duals $\phi^\alpha = \sum \left(S^{-1}\right)^{\alpha\beta}\phi_\beta$, which have a much larger localisation region than the NGWFs. Thus, even if the orbitals are localised, the kernel is not automatically sparse; however, based on the near-sightedness principle, one can enforce sparsity by assuming that if two NGWFs are far apart, their contribution to the density kernel is null.

The locality of the NGWFs and the truncation of the density-kernel ensure that the information in the density matrix actually scales as $O(N)$ with the system size for insulating or semiconducting systems. However, we did not employ kernel truncation, as this is usually required only for systems much larger than ours.

In order to obtain accurate results, self-consistency is still required; in ONETEP this is done by optimising both the density-kernel (inner loop [60]) and NGWFs (outer loop [58]). This method avoids charge sloshing and is equivalent to Ensemble DFT in plane-wave codes, where one optimises state occupancies.

As previously explained, the density kernel must be idempotent, but the cost of explicitly imposing Eq. 2.1.6 scales as $O(N^3)$ for dense matrices. Since ONETEP needs to be linear-scaling, an alternative solution was required: the LNV method [61, 62], which ensures implicit idempotency at the ground state. This implies that the density-kernel is expressed in terms of an auxiliary matrix L:

$$\mathbf{K} = 3\mathbf{LSL} - 2\mathbf{LSLSL} \ , \tag{2.1.7}$$

where L is optimised iteratively in order to minimise the energy functional. This method has the advantage that the energy functional does not have multiple minima (being cubic as a function of L).

2.1.3 NGWF formalism

In ONETEP, one utilises only a small number of dynamic (iteratively optimised) local orbitals [1], which should have the same accuracy as a large fixed basis set. Therefore, the previously defined NGWFs are optimised to adapt to their chemical environment, ensuring transferability, and eliminating the need for basis-set superposition error (BSSE [63]) corrections [64]. In the following, we will present some defining properties of NGWFs.

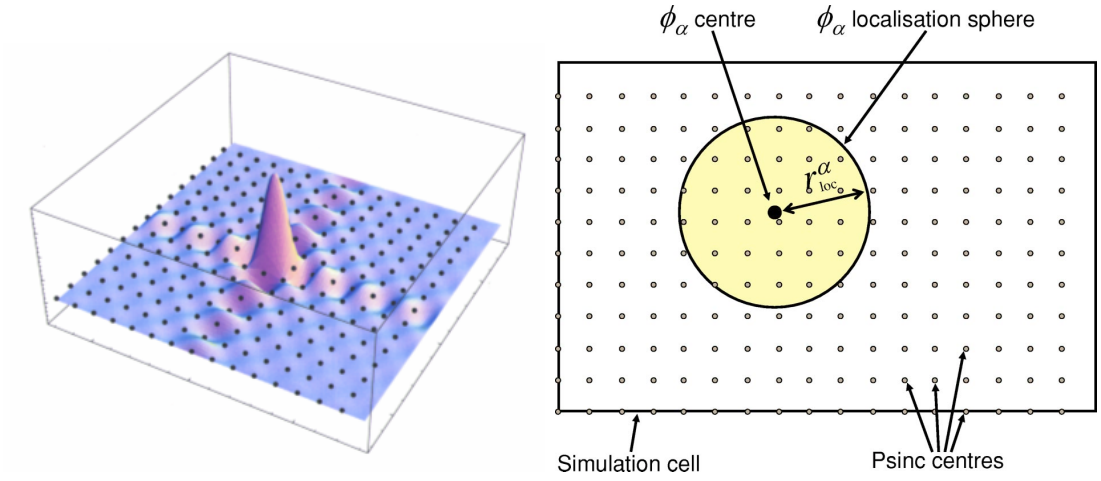


Figure 2.2: Left: 2D psinc centered at a grid point in the simulation cell (courtesy of Haynes et al. [64]). Right: NGWF expansion in terms of the psinc basis (courtesy of Skylaris et al. [1]).

The underlying basis in ONETEP comprises psinc functions centred on the points of a regular real-space grid [1], i.e. periodic bandwidth-limited delta functions (Fig. 2.2 - left inset):

$$D_i(\mathbf{r}) = \frac{1}{N}\sum_p e^{i\mathbf{k}_p\cdot(\mathbf{r}-\mathbf{r}_i)} \ , \tag{2.1.8}$$

with N being the number of points in the simulation cell, $\mathbf{r}_i$ is a grid point in the real-space simulation cell and $\mathbf{k}_p$ is a reciprocal space point, consisting of multiples of the simulation cell reciprocal lattice vectors. The maximum value of $\mathbf{k}_p$ is controlled by the

kinetic energy cutoff. The psinc functions have some useful properties: they are non-zero only on the grid point on which they are centred, they share the periodicity of the cell, and are mutually orthogonal. The NGWFs are expanded in terms of this psinc basis (Fig. 2.2 - right inset):

$$\phi_\alpha(\mathbf{r}) = \sum_{i \in R_\alpha} D_i(\mathbf{r}) c^i{}_\alpha \ , \tag{2.1.9}$$

where R_α is the localisation radius of the NGWF and $c^i{}_\alpha$ is the coefficient of the NGWF $\phi_\alpha(\mathbf{r})$ in terms of the psinc at the grid point i.

In ONETEP, the computational effort is further reduced by performing Fast Fourier Transforms (FFTs) on so-called FFT boxes instead of the whole simulation cell. These boxes are centred on the NGWFs and they encompass all the overlapping neighbours of the respective NGWF (i.e. the box has a lateral size six times larger than the NGWF cutoff radius). Therefore, linear scaling is maintained, as the FFT domain only depends on the radius of the NGWFs, making the number of FFT operations scale linearly with the number of NGWFs, i.e. the system size.

2.2 Projector Augmented Wave (PAW) formalism

In DFT, valence wavefunctions are highly irregular near the ion core, as they are need to always be orthogonal to the core wavefunctions. This lack of smoothness implies the need of a large number of Fourier coefficients for an accurate representation. However, the crucial physical interactions, such as chemical bonding, are mostly related to the smooth tails of the valence wavefunctions. This opens up the possibility of using smooth pseudised (PS) wavefunctions, thus replacing the all-electron (AE) eigenstates and potentials, such that the required computational effort is reduced, with little effect on the calculation accuracy. However, such a transformation should only affect the wavefunction near the ion core, as the AE and PS wavefunctions and potentials need to identical after a predetermined cut-off radius, in order to accurately represent bonding.

While there are several variations of this approach, the most promising is the Projector-Augmented Wave formalism (PAW), introduced in 1994 by Blöchl [50]. This framework is the theoretical generalisation of the pseudopotential concept, and it allows for smoother

PS wavefunctions due to the lack of constraints regarding norm conservation. Lastly, one can obtain all-electron quantities from the PS equivalents.

2.2.1 Wavefunctions

Such that the pseudisation leaves the wavefunction tails unchanged, one defines a transformation that acts only inside a certain spherical augmentation region around the atom:

$$|\Psi_n\rangle = \hat{\tau}\,|\widetilde{\Psi}_n\rangle = \left[1 + \sum_i \Big(|\varphi_i\rangle - |\widetilde{\varphi}_i\rangle \Big) \langle \widetilde{p}_i| \right] |\widetilde{\Psi}_n\rangle \ , \tag{2.2.1}$$

with $\hat{\tau}$ being the transformation operator between the PS $|\widetilde{\Psi}_n\rangle$ and AE $|\Psi_n\rangle$ eigenstates (all PS quantities have "$\sim$" as accent). The PS wavefunctions are expanded into a basis of PS partial waves $|\widetilde{\varphi}_i\rangle$:

$$|\widetilde{\Psi}_n\rangle = \sum_i |\widetilde{\varphi}_i\rangle \, \langle \widetilde{p}^i | \widetilde{\Psi}_n \rangle \ , \tag{2.2.2}$$

where the overlap between the wavefunctions and projectors $|\widetilde{p}^i\rangle$ acts as coefficients; an equivalent relation applies to the AE wavefunctions, using AE partial waves $|\varphi_i\rangle$. The partial waves obey the same transformation as the wavefunctions (Eq. 2.2.1), with the additional property that they are restricted to be zero outside the augmentation sphere:

$$|\varphi_i\rangle = \hat{\tau}\,|\widetilde{\varphi}_i\rangle \tag{2.2.3}$$

PAW datasets usually contain the radial parts of the partial waves and projectors, but one also needs to consider a spherical component, according to their quantum numbers:

$$\varphi_i(\mathbf{r}) = \sum_{n_i, l_i, m_i} \frac{\varphi_{n_i,l_i}(r)}{r} S_{l_i,m_i}(\hat{\mathbf{r}}) \ , \tag{2.2.4}$$

where the same relation applies to PS partial waves or projectors, n_i, l_i and m_i are quantum numbers, and S_{l_i,m_i} is a real spherical harmonic.

Lastly, since the PS wavefunctions are not orthogonal in the same sense that the AE eigenstates are ($\langle \Psi_i | \Psi_j \rangle = \delta_{ij}$), one needs to redefine the orthogonality relation through a new metric:

$$\langle \Psi_i | \Psi_j \rangle = \langle \widetilde{\Psi}_i | \underbrace{\hat{\tau}^\dagger \hat{\tau}}_{\hat{S}} \widetilde{\Psi}_j \rangle = \delta_{ij} \ , \tag{2.2.5}$$

called an augmented overlap operator $\hat{S}$:

$$\hat{S} = 1 + \sum_{a,b} |\widetilde{p}^a\rangle \underbrace{\left[\langle \varphi_a | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \widetilde{\varphi}_b \rangle \right]}_{O_{ab}} \langle \widetilde{p}^b | \tag{2.2.6}$$

2.2.2 Densities

In an equivalent manner to the previously described wave-function pseudisation procedure, the all-electron charge density can be viewed as a PS density on the whole simulation grid, from which one removes the sphere of PS density (in the augmentation region) and replaces it with a sphere of AE density, as also shown in Fig. 2.3:

$$\rho(\mathbf{r}) = \underbrace{\sum_n f_n \langle \widetilde{\Psi}_n | \mathbf{r} \rangle \langle \mathbf{r} | \widetilde{\Psi}_n \rangle}_{\widetilde{\rho}(\mathbf{r})} + \underbrace{\sum_{i,j} \rho^{ij} \langle \varphi_i | \mathbf{r} \rangle \langle \mathbf{r} | \varphi_j \rangle}_{\rho^1(\mathbf{r})} - \underbrace{\sum_{i,j} \rho^{ij} \langle \widetilde{\varphi}_i | \mathbf{r} \rangle \langle \mathbf{r} | \widetilde{\varphi}_j \rangle}_{\widetilde{\rho}^1(\mathbf{r})} \ , \tag{2.2.7}$$

where we have defined a projector density kernel ρ^{ij} (f_n are the occupation numbers):

$$\rho^{ij} = \sum_n \langle \widetilde{\Psi}_n | \widetilde{p}^i \rangle \, f_n \, \langle \widetilde{p}^j | \widetilde{\Psi}_n \rangle \ , \tag{2.2.8}$$

As an added constraint, the PS augmentation density ($\widetilde{\rho}^1$) needs to have the same mul-

Figure 2.3: Schematic of augmentation and pseudisation procedure for densities.

tipole moment as the AE augmentation density (ρ^1). Therefore, a compensation density $\hat{\rho}(\mathbf{r})$, non-zero only inside the augmentation region, is added to the pseudo-densities ($\widetilde{\rho}(\mathbf{r})$ and $\widetilde{\rho}^1(\mathbf{r})$):

$$\hat{\rho}(\mathbf{r}) = \sum_{i,j} \rho^{ij} \hat{Q}_{ij} \ , (\mathbf{r}) \tag{2.2.9}$$

where Q_{ij} are multipole moments. Therefore one maintains the charge density quadrature, while allowing the correct description of the multipole. Lastly, PAW datasets also include the PS ($\widetilde{\rho}_c$) and AE (ρ_c) core-electron charges, which are important for the exchange-correlation energy; by also adding the ion charge, one obtains $\widetilde{\rho}_{Z_c}$ and ρ_{Z_c}, which are used in the definition of the local pseudo-potential.

2.2.3 Energy and pseudo-Hamiltonian

One quantity that will be the starting point of our discussion in Chapter 4 is the PAW electronic energy:

$$
\begin{aligned}
E^{\mathrm{el}}_{\mathrm{PAW}} = &\sum_{i}^{N_e/2} f_i \left\langle \widetilde{\psi}_i \right| -\frac{1}{2}\nabla^2 + v_{\mathrm{H}}\left[\widetilde{\rho}_{Z_c}\right] \left| \widetilde{\psi}_i \right\rangle + E_{\mathrm{H}}\left[\widetilde{\rho}+\hat{\rho}\right] + E_{\mathrm{XC}}\left[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c\right] + \\
&\sum_{a,b} \rho^{ab} \Bigg\{ \int v_{\mathrm{H}}\left[\widetilde{\rho}_{Z_c}\right] \hat{Q}_{ab}\ \mathrm{d}\mathbf{r} + \langle \varphi_a | -\frac{1}{2}\nabla^2 + v_{\mathrm{H}}\left[\rho_{Z_c}\right] | \varphi_b \rangle - \\
&\langle \widetilde{\varphi}_a | -\frac{1}{2}\nabla^2 + v_{\mathrm{H}}\left[\widetilde{\rho}_{Z_c}\right] | \widetilde{\varphi}_b \rangle - \int_{\Omega} v_{\mathrm{H}}\left[\widetilde{\rho}_{Z_c}\right] \hat{Q}_{ab}\ \mathrm{d}\mathbf{r} \Bigg\} + \\
&E_{\mathrm{H}}\Big|_{\Omega}\left[\rho^1\right] + E_{\mathrm{XC}}\Big|_{\Omega}\left[\rho^1+\rho_c\right] - E_{\mathrm{H}}\Big|_{\Omega}\left[\widetilde{\rho}^1+\hat{\rho}\right] - E_{\mathrm{XC}}\Big|_{\Omega}\left[\widetilde{\rho}^1+\hat{\rho}+\widetilde{\rho}_c\right] = \\
&\sum_{i}^{N_e/2} f_i \langle \widetilde{\psi}_i | \widetilde{H} | \widetilde{\psi}_i \rangle - E_{dc}[\rho]
\end{aligned}
\tag{2.2.10}
$$

where "H" and "XC" denote Hartree and exchange-correlation energies, respectively. Energies with Ω are calculated in the augmentation sphere, while the first line of the previous equation is computed on the entire simulation grid. In the last equality we have shown the standard decomposition of the energy into a bandstructure term and a density-dependent double-counting element. The former component is determined by a Kohn-Sham the pseudo-Hamiltonian $\widetilde{H}$, containing an effective potential $\widetilde{v}_{\mathrm{eff}}$ which describes the pseudo-, Hartree, and exchange-correlation potentials:

$$
\begin{aligned}
\widetilde{H} &= -\frac{1}{2}\nabla^2 + \widetilde{v}_{\mathrm{eff}} + \sum_{a,b} |\widetilde{p}^a\rangle \left[\hat{D}_{ab} + D^1_{ab} - \widetilde{D}^1_{ab}\right] \langle \widetilde{p}^b | \\
\widetilde{v}_{\mathrm{eff}}(\mathbf{r}) &= v_{\mathrm{H}}[\widetilde{\rho}_{Z_c}](\mathbf{r}) + v_{\mathrm{H}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}) + v_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})
\end{aligned}
\tag{2.2.11}
$$

The remaining non-local component of the pseudo-Hamiltonian is described by

$$
\begin{aligned}
\hat{D}_{ab} &= \int \widetilde{v}_{\mathrm{eff}}(\mathbf{r})\ \hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r} \\
D^1_{ab} &= \langle \varphi_a | -\frac{1}{2}\nabla^2 + v^1_{\mathrm{eff}} | \varphi_b \rangle \\
\widetilde{D}^1_{ab} &= \langle \widetilde{\varphi}_a | -\frac{1}{2}\nabla^2 + \widetilde{v}^1_{\mathrm{eff}} | \widetilde{\varphi}_b \rangle + \int_{\Omega} \widetilde{v}^1_{\mathrm{eff}}(\mathbf{r})\ \hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r}\ ,
\end{aligned}
\tag{2.2.12}
$$

where v^1_{eff} and $\widetilde{v}^1_{\mathrm{eff}}$ are all-electron and pseudised equivalents of $\widetilde{v}_{\mathrm{eff}}$ on the augmentation-sphere. More nonessential details regarding the general PAW framework can be found in the ABINIT plane-wave implementation [65] of PAW.

Lastly, for the particularities of PAW in LS-DFT, we refer the reader to the detailed ONETEP implementation as written by Nicholas Hine [66].

2.3 Spectral function unfolding

In this thesis, especially in Chapter 3, we have used the concept of bandstructure unfolding, which has already been described in our previous work [36] and my Master thesis [37]. However, for the convenience of the readers we briefly revisit the theory.

An extensively used experimental technique to probe the electronic structure of materials is angular resolved photoelectron spectroscopy (ARPES), which registers the photocurrent [67, 68] induced by photo-excitations:

$$I(\mathbf{k},\omega) \propto f(\omega)|M_{f,i}(\mathbf{k})|^2 A_{\mathbf{k}j,\mathbf{k}j}(\omega) \ , \tag{2.3.1}$$

Neglecting the polarization-dependent dipole matrix $|M_{f,i}(\mathbf{k})|^2$, we see that this current is directly related to the one-particle Kohn-Sham spectral function $A_{\mathbf{k}j,\mathbf{k}j}(\omega)$. The latter represents probability that an electron of momentum $\mathbf{k}$ can be added/removed into/from band j with an energy of ω, and its operator is defined as:

$$\hat{A}(\omega) = -\frac{1}{\pi}\mathrm{Im}\sum_{\mathbf{K},J}\frac{|\Psi_{\mathbf{K}J}\rangle\langle\Psi_{\mathbf{K}J}|}{\omega+i\eta-\epsilon_{\mathbf{K}J}} = \frac{\eta}{\pi}\sum_{\mathbf{K},J}\frac{|\Psi_{\mathbf{K}J}\rangle\langle\Psi_{\mathbf{K}J}|}{(\omega-\epsilon_{\mathbf{K}J})^2+\eta^2} \ , \tag{2.3.2}$$

with the eigenvector and eigenvalue of the system state being $|\Psi_{\mathbf{K}J}\rangle$ and $\epsilon_{\mathbf{K}J}$, while η is a small broadening factor.

The bandstructure is not a useful concept in the supercells we have simulated, as one cannot envision how electronic effects relate back to the Brillouin zone (BZ) of each monolayer's unit cell. To solve this problem, we project the supercell spectral function onto the subsystem (monolayer) of interest and unfold it to the unit cell BZ:

$$A^I_{\mathbf{k}j,\mathbf{k}j} = \sum_{\substack{\mathbf{K},\\ \rho\in I}}\langle\Psi_{\mathbf{k}j}|\Psi_{\mathbf{K}J}\rangle\, A_{\mathbf{K}J,\mathbf{K}J}\,\langle\Psi_{\mathbf{K}J}|\phi^\rho\rangle\langle\phi_\rho|\Psi_{\mathbf{k}j}\rangle \ , \tag{2.3.3}$$

where $A^I_{\mathbf{k}j,\mathbf{k}j}$ is the spectral function projected on layer I in the representation of the primitive cell eigenstates $|\Psi_{\mathbf{k}j}\rangle$, $A_{\mathbf{K}J,\mathbf{K}J}$ is the heterostructure spectral function for the supercell eigenstates $|\Psi_{\mathbf{K}J}\rangle$ and $\sum_{\rho\in I}|\phi^\rho\rangle\langle\phi_\rho|$ projects onto the atomic orbitals of subsystem I.

Chapter 3

Two-dimensional heterostructures

3.1 hBN/BP

In this section we describe our work related to the interfacing of black phosphorus (BP), a semiconducting layered material, with hexagonal boron nitride (hBN), an insulating counterpart. We remind the reader that our main goal was to prove that hBN can safely shield BP from external interactions, limit its band gap decrease in multilayers, as well as enable its use as a tunnelling field effect transistor. The work presented here has been published as Nano Letters 16, 4, 2586-2594 (2016) [69].

3.1.1 Methodology

Before performing any calculations, we were required to pinpoint the methodological details, such as structural parameters for the considered BP and hBN layers, the kinetic-energy cutoff, and the optimal size of the supercells in ONETEP.

The exchange-correlation functional was chosen to be optB88-vdW, since previous studies [36] have shown that its ab-initio description of the van-der-Waals interaction predicts the correct structure and binding energies in weakly bound layers. We have employed the PAW formalism for most of the ONETEP calculations that will be mentioned from this point onwards, in order to have access to all-electron quantities in a cost-effective manner. We chose the PAW datasets developed by Garrity et al [70], as they have been extensively validated for elemental structures and simple covalent compounds, and have proven effective in our previous study on $MoS_2/MoSe_2$ heterostructures[36]. The P and N atoms contain 5 valence electrons ($2s^2, 2p^3$ for N / $3s^2, 3p^3$ for P), while B has only 3 valence electrons ($2s^2, 2p^1$).

Since the initial convergence tests were performed using Quantum Espresso[71], we also use the equivalent ultra-soft pseudopotential (USPP) versions of these PAW datasets.

While requiring a higher charge density cutoff, Garrity et al. [70] showed that USPPs produce essentially identical results to the PAW datasets.

First, we have determined the structures of the considered materials. Few-layered hexagonal boron nitride (hBN) was synthesised at the same time as graphene [2]. Its structure is rather simple, having a hexagonal symmetry (space group $P6_3/mmc$), with one B and one N atom per each periodic unit, each occupying a different sub-lattice site. The monolayer structure is shown in Fig. 3.1.a; in an previous paper, I have shown that the bilayer form is most stable in the AA' stacking form [72], where the B in one layer eclipses the N in the other layer, and vice versa (Fig. 3.1.b). The BP monolayer, however, has a more complicated structure. Its puckered orthorhombic structure (space group D_{2h}^{18}) contains two subplanes of P atoms, with each BP monolayer unit containing 4 atoms (Fig. 3.1.d). For the bilayer, the AB stacking (Fig. 3.1.e) is the most stable [73], but in this study we will also be interested in the slightly less stable AA stacking (Fig. 3.1.f). The latter configuration is conceptually easier to envision, and the BP stacking type does not change our qualitative conclusions.

For both hBN and BP, the in-plane lattice vectors have been denoted as a, b, the out of plane vector as c (relevant only for the bulk structure), and the interlayer distance as d (relevant for the bilayer cases). In the case of bilayer BP, d is redefined as the distance between the lower P subplane of the top monolayer and the upper P subplane of the bottom monolayer (Fig. 3.1.d)

For optimising the structures of each separate material, we have used small unit cells and the QuantumEspresso [71] plane-wave DFT code. In the following, all the geometry optimisation were run at a large $1,000$ eV kinetic energy cutoff, with a $10,000$ eV charge-density cutoff. The force tolerance was set to 10^{-4} a.u./bohr ($\approx 5 \cdot 10^{-3}$ eV/Å) in all cases except for the phosphorene bilayers, which could not be converged beyond $5 \cdot 10^{-4}$ a.u./bohr ($\approx 2.5 \cdot 10^{-2}$ eV/Å); the stress threshold was set to 0.25 kbar. The Brillouin zone was sampled by a 12×12×1 k-point mesh for hBN and a 13×10×1 one for BP; in the bulk structures we have also considered 4 out-of-plane k-points. For the monolayers and bilayers, the simulation cell was chosen to have a height of 30 Å, in order to avoid interactions between periodic images.

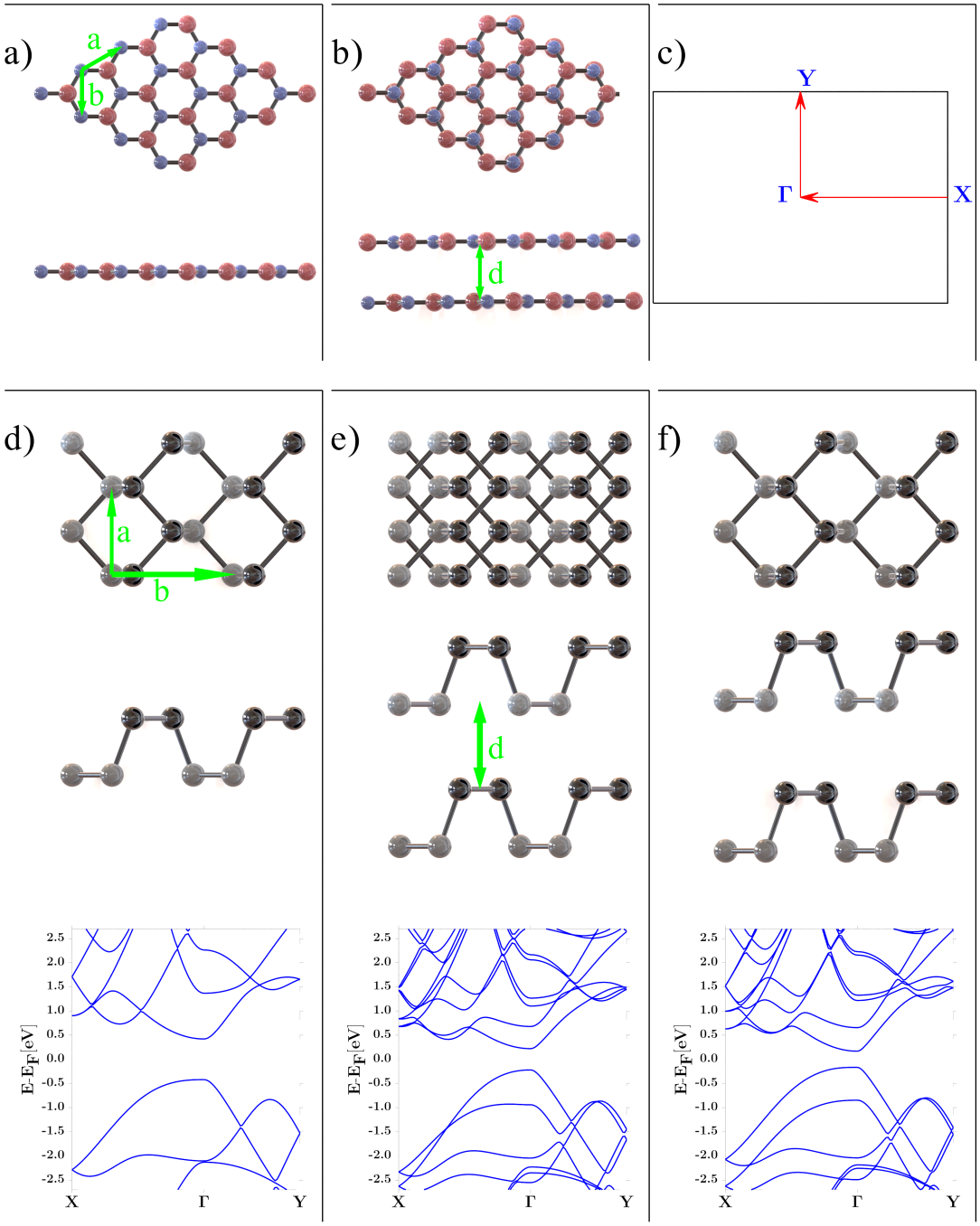


Figure 3.1: Top and side view of a 4x4 supercell of the hBN monolayer (a), and bilayer in AA' stacking (b); the red(blue) spheres are N(B) atoms. Brillouin zone for the BP monolayer/bilayer (c). Top and side view of a 2x2 supercell, as well as band-structure plots, for the BP monolayer (d), bilayer in AB-stacking (e) and AA-stacking (f). For clarity, the P atoms in the lower(upper) sub-plane of each monolayer are grey(black). In all cases, a, b are the in-plane lattice vectors, and d is the interlayer distance. The reference for the band-structure energy scale is the mid-gap level.

We first optimised the lattice parameters of the bulk structures (see Table 3.1), as uncontroversial experimental data was already available for both hBN and BP in bulk form. Note that the BP bulk has AB stacking (Fig. 3.1.e), while the hBN bulk has AA' stacking (Fig. 3.1.b).

Table 3.1: Calculated lattice parameters (in Å) for bulk BP and hBN, experimental values (EXP.) and other theoretical (THEO.) DFT or higher-order results.

	bulk hBN	EXP. bulk hBN [74]	THEO. bulk hBN [72]	bulk BP	EXP. bulk BP [75]	THEO. bulk BP [73]
a [Å]	2.51	2.50	2.49	3.36	3.31	3.34
b [Å]	2.51	2.50	2.49	4.48	4.38	4.48
c [Å]	6.60	6.66	6.68	10.75	10.48	10.73

The results are a testament to the behaviour of the optB88-vdW functional: for hBN, our results almost perfectly agree with the experimental ones, with a maximum relative error of $\approx 1\%$, while for BP the difference is $< 3\%$. With such encouraging results for the bulk structures, we have also optimised the monolayers and bilayers, as shown in Table 3.2.

Table 3.2: Optimised lattice parameters (in Å) for BP and hBN monolayer (mono) and bilayers (bi-"stacking").

	mono hBN	bi-AA' hBN	mono BP	bi-AB BP	bi-AA BP
a [Å]	2.51	2.50	3.34	3.35	3.34
b [Å]	2.51	2.50	4.59	4.54	4.53
d [Å]		3.33		3.20	3.49

The lattice parameters for hBN match both experiment [76] and higher-order MP2 theory [72]. While experimental structural parameters for BP monolayers and bilayers could not be found, our results agree perfectly with other DFT works employing the optB88-vdW [73], PBE [24, 47], vdW-DF2 and PBE-D2 [47] functionals. The interlayer distance for the bilayer BP in the AB stacking is 0.29 Å smaller than for the AA stacking, in perfect agreement with Dai et al. [73]. Consequently, we shall use all these converged values in constructing the BP/hBN heterostructures (see section 3.1.3).

Some explanations are necessary regarding the bandstructures of the monolayer and bilayer BP (bottom insets of Fig. 3.1.d,e,f). For clarity, the orthorhombic Brillouin zone of BP is shown in Fig. 3.1.c. As we have mentioned in the introduction, there is a significant quantum confinement effect upon the stacking of the BP layers: DFT predicts a significant band gap reduction from 0.84 eV in the monolayer case, to 0.44 eV and 0.33 eV for the AB and AA stacking of the bilayer, respectively. These values are in agreement with others [14], although unsurprisingly underestimated with respect to experiment [15], hybrid-DFT [73, 77] and GW+BSE [14] calculations. In all the cases, the anisotropy of the hole effective mass is obvious, as the mass in the zig-zag direction (3.08 m_0 along $\Gamma - X$) is much higher than the one in the armchair direction (0.16 m_0 along $\Gamma - Y$), in agreement with experimental studies [17, 18, 78, 20]. While structural differences between the AA and AB stackings for BP bilayers are obvious (Fig. 3.1.e,f), the bandstructures are almost identical, the most major differences being a reduction of the band gap in the AA-case by 0.11 eV and a slightly lower hole effective mass along $\Gamma - X$ for the AB stacking. This obvious similarity between the electronic band-structures of the AA and AB alignments has proven useful in our choice of heterostructures construction (Sec. 3.1.3).

In QuantumEspresso, the kinetic energy (KE) cutoff controls the number of plane waves used to describe the Kohn-Sham orbitals, in a similar manner to how the KE cutoff fixes the psinc grid spacing in ONETEP. Thus, the minimal KE cutoff required for sufficient convergence in QuantumEspresso can be assumed to be adequate for any equivalent ONETEP calculation [79].

For this test, the convergence criterion was taken to be the interlayer binding energy of a hBN AA'-bilayer (Fig. 3.1.b) and a BP AB-bilayer (Fig. 3.1.d), respectively. Apart from the KE cutoff, we have maintained all the previously discussed parameters. The results are shown in Table 3.3.

It is clear that the values are well converged ($\ll$ 1.0 meV/bilayer-unit) even for small kinetic energy cutoffs. However, to ensure a sufficient sampling of the NGWFs in the ONETEP calculations, we have chosen a reasonable value of 800 eV for the KE cutoff.

Table 3.3: Convergence test for the KE cutoff. Binding energy (meV/bilayer-unit) for the hBN AA'-bilayer and the BP AB-bilayer.

K.E. [eV]	hBN ΔE [meV]	BP ΔE [meV]
500	-124.24	-441.94
600	-124.10	-441.99
700	-124.10	-442.01
800	-124.06	-442.04
900	-124.05	-442.05
1000	-124.05	-442.06

Another important parameter that needs to be converged is the k-point sampling in QuantumEspresso. In the case of ONETEP, this is equally important: as ONETEP only considers the Γ point in reciprocal space, the k-point sampling in traditional DFT is equivalent to the supercell size in ONETEP. Therefore, the minimum required k-point sampling was determined by once again performing a convergence test in QuantumEspresso, with the binding energies for the hBN and BP bilayers being the convergence criterion. The results are shown in Table 3.4:

Table 3.4: Convergence test for in-plane k-point sampling. Binding energy (meV/bilayer-unit) for bilayered hBN (AA' stacking) and BP (AB stacking).

sampling	hBN ΔE [meV]	sampling	BP ΔE [meV]
6x6	-123.98	8x6	-442.65
9x9	-124.04	11x8	-442.14
12x12	-124.05	13x10	-442.06
15x15	-124.05	16x12	-442.06

Clearly, all the considered cases provide sufficient convergence, even for sparse k-point sampling. We have therefore chosen the minimum k-point sampling to be 12×12 for hBN (19.95 points Å^2) and 11×8 (34.16 points Å^2) for BP. This was equivalent to the condition that the hBN/BP heterostructures need to contain at least a 11x8 supercell of BP and a 12x12 supercell of hBN.

3.1.2 NGWF selection

In order to study the hBN/BP heterostructures in ONETEP, we had to transition from the QuantumEspresso convergence tests to the parameters utilised in the ONETEP calculation.

Thus, the remaining task was to determine the minimum NGWF radius that would ensure sufficient accuracy for our calculations. The convergence tests from the previous subsections allow us to employ a kinetic energy cutoff of 800 eV, and as test system we once again chose the hBN AA'-bilayer and the BP AB-bilayer, but now in the form of a 12x12 supercell for hBN and a 11x8 supercell for BP. Once again, the simulation cell height was set to 30.0 Å in order to avoid interaction between periodic images.

First, we needed to fix the number of NGWFs. As explained in the beginning of the chapter, the PAW sets consider 5 valence electrons for N ($2s^2, 2p^3$), 3 valence electrons for B ($2s^2, 2p^1$) and 5 valence electrons for P ($3s^2, 3p^3$). In general we need one NGWF to describe the s-shell, three for the p-shell and five for the d-shell. Thus, the minimum number of NGWFs for all three atom types is 4, since only s and p shells contribute. However, upon testing we found that the NGWFs on P atoms require additional polarisation for the calculation to converge, meaning that we also include d-shell NGWFs for P. In the end, we use 4 NGWFs for B and N, and 9 for P.

By default, the NGWFs are considered converged when the RMS value for the NGWF gradient is smaller than $2 \cdot 10^{-6}$. While this proved sufficient for hBN, the BP bilayer required a stricter value ($1.5 \cdot 10^{-6}$) for sufficient accuracy. Moreover, BP also proved more pretentious regarding the fine-grid sampling, on which the density and potentials are mapped: while a factor of 2 between the standard-grid (defined by the kinetic-energy cutoff) and the fine-grid was sufficient for hBN, BP required a factor of 4. With all the parameters in place, we calculated the binding energy as a function of NGWF radius (Table 3.5):

Once again, the hBN binding energy was well-converged even for small values of the NGWF radius. However, in the BP case, sufficient convergence (energy change of ≤ 1.0 meV/bilayer-unit) was achieved for a minimum radius of 11.0 bohr.

Table 3.5: Convergence test for NGWF localisation radius (r). Binding energy (meV/bilayer-unit) for the hBN AA'-bilayer and the BP AB-bilayer.

r [a.u.]	hBN ΔE [meV]	BP ΔE [meV]
9.0	-124.30	-452.20
10.0	-124.98	-450.00
11.0	-125.18	-445.19
12.0	-125.34	-444.82

In conclusion, our calculations on hBN/BP heterostructures utilise NGWFs with a radius of 11.0 bohr for both hBN and BP. The other parameters will be set according to the results of our previous convergence tests: 800 eV kinetic energy cutoff, a fine grid scale factor of 4.0, and a minimum supercell size of 11x8 for BP and 12x12 for hBN.

3.1.3 Heterostructure construction

We also present the construction technique of the hBN/BP heterostructures. Not only do these materials have different periodicity patterns (hexagonal for hBN and orthorhombic for BP), but also significantly different lattice constants, meaning that their superposition will produce Moiré patterns. Thus, for our purposes we have adopted the work of K. Hermann [80], who describes construction techniques for coincidence cells of such Moiré patterns.

Since it is easier to obtain coincidence cells for materials belonging to the same crystal family, we have created an orthorhombic unit cell for hBN, containing two primitive hBN units; afterwards we could simply follow the instructions of K. Hermann seamlessly [80]. We had to maintain the supercell-size constraints determined in Sec. 3.1.1: each BP monolayer needs to have at least 11x8 units, while each hBN monolayer needs to contain at least 288 atoms (i.e. 12x12 units).

To limit the size of the simulation cells, we have adopted two strategies: first, we chose specific values of the rotation angle between the hBN and BP layers which keep the supercell size relatively small. Secondly, we allowed a small strain in the BP layers, i.e. 1.0% in both the armchair and zig-zag direction, a value smaller than the 2.0% required to noticeably modify the BP electronic structure [24]. We chose to strain the monolayer

BP (41.3 GPa and 106.4 GPa in the armchair and zig-zag directions, respectively [81]) instead of the monolayer hBN (716–977 GPa [82]) because the smaller elastic modulus of the former would make it more prone to in-plane deformations.

The resulting hBN/BP/hBN and BP/hBN/BP heterostructures [69] are shown in Fig. 3.2.f and 3.2.g, respectively. Both of them have identical lattice constants (i.e. $R_1 = 39.90$ Å and $R_2 = 46.07$ Å) and a rotation angle of 40.89° between the hBN and BP layers. Each BP layer contains 12x10 primitive unit cells (i.e. 480 atoms) and each hBN layer has 672 atoms; thus, the entire hBN/BP/hBN supercell has 1824 atoms, while the BP/hBN/BP one has 1632 atoms, meaning that it is impractical to simulate such interfaces with traditional plane-wave DFT.

In both heterostructures, the outlying layers are identical, meaning that in the BP/hBN/BP heterostructure the BP layers are actually in the AA stacking (Fig. 3.1.f), instead of the slightly more stable AB stacking (Fig. 3.1.e). However, using the AA stacking reduces the complexity of the tunnelling calculations outlined in section 3.1.4. Also, the band-structures in the AA and AB configurations of BP are very similar (Sec. 3.1.1); since our goal was to decouple the BP layers with a hBN spacer, our conclusions are transferable to any initial BP stacking.

Lastly, no other work considers using hBN as a spacer between BP sheets at the time of writing this thesis, but there are several papers addressing hBN/BP/hBN heterostructures [83, 84]. In the aforementioned sources, the authors have used a rather large (3.0%) strain in BP, thus obtaining much smaller supercells, with a 0° interlayer rotation. However, too much strain can affect the electronic properties of BP [24], so we believe that a value of 3.0% is significant. Furthermore, the relative rotation between hBN and BP in our structures is useful for future comparisons between interfaces with and without interlayer twist.

3.1.4 Results and discussion

We have used the computational setup described in the end of section 3.1.2 to perform a relaxation of the atomic positions for the hBN/BP/hBN and BP/hBN/BP interfaces in

ONETEP. The only difference is the simulation cell height, which we have set at a higher value of 40.0 Å, since the heterostructures are thicker than the bilayers.

The force tolerances were set to $2 \cdot 10^{-3}$ a.u./bohr ($\approx 10^{-2}$ eV/Å), an acceptable value for such large structures. For BP/hBN/BP, the default NGWF RMS (root mean square) gradient threshold ($2.0 \cdot 10^{-6}$) and an energy tolerance of $3.0 \cdot 10^{-7}$ a.u./atom were sufficient to achieve an accurately converged result; however, for the hBN/BP/hBN structure, we had to use tighter tolerances: $1.5 \cdot 10^{-6}$ for the NGWF RMS gradient and $2.0 \cdot 10^{-7}$ a.u./atom for the energy tolerance. Both calculations were highly computationally demanding, due to the large number of atoms. We also needed the optimised structures of the monolayer BP and bilayer BP in AA stacking; however, by simply constructing 11x8 supercells (in agreement with our convergence tests in section 3.1.1) of the primitive units we had previously optimised in QuantumEspresso, the force tolerances were directly below the selected threshold.

In the optimised hBN/BP/hBN, the distance between the top hBN layer and the top P subplane of BP was $d_1 = 3.54 \pm 0.03$ Å, while the distance between the bottom hBN layer and the bottom P subplane was $d_2 = 3.58 \pm 0.02$ Å. These values were within the range defined by the values of Cai et al.[83] (3.46 Å) and Rivero et al. [84] (3.68 Å), both being for hBN/BP heterostructures without interlayer rotation. The BP layer also remained rather rigid, with the corrugation of the P atoms being < 0.05 Å throughout the heterostructure.

However, for the BP/hBN/BP structure, the distance between the bottom subplane of the top BP layer and the hBN sheet was $d_1 = 3.39 \pm 0.12$ Å, while the distance between the top subplane of the bottom BP layer and the hBN sheet was $d_2 = 3.60 \pm 0.12$ Å. Thus, the corrugation of the hBN layer is much more significant, being ≈ 0.24 Å across the unit cell. The BP once again remains rather rigid, with the interlayer distance between the BP layers varying by only 0.10 Å.

One of the main goals of this work [69] was to observe if any significant alterations occur to the electronic-structure of BP when it is either encapsulated by hBN or when hBN is used as a spacer. As a measure of such effects, we have utilised the one-particle Kohn-Sham spectral function $A_{\mathbf{k}j,\mathbf{k}j}(\omega)$, the concept of which has been presented in Sec. 2.3.

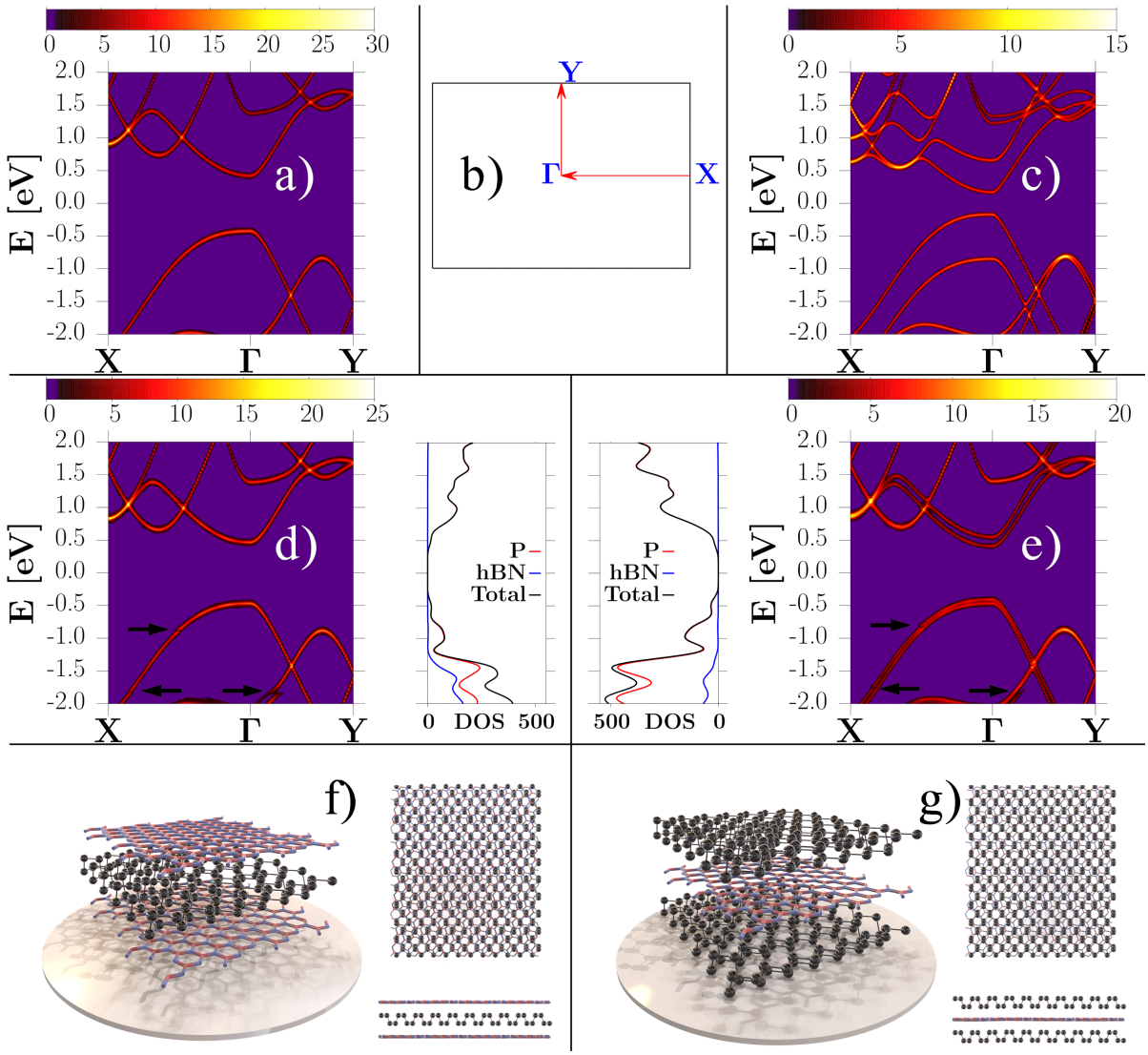


Figure 3.2: Unfolded spectral function for the BP monolayer (a), the top BP sheet in the BP AA-stacked bilayer (c), the BP layer in the hBN/BP/hBN interface (d), and the top BP sheet in the BP/hBN/BP interface (e). Primitive-cell Brillouin-zone sampling path depicted in (b). Local densities of states of BP and hBN shown for hBN/BP/hBN (d) and BP/hBN/BP (e). Energy scales are with respect to the mid-gap level. Black arrows in (d),(e) highlight band discontinuities due to hBN-BP interactions. Schematic representation, and top/side views of hBN/BP/hBN (f) and BP/hBN/BP (g) – P atoms are black spheres.

Bandstructure effects

Firstly, we investigated whether hBN encapsulation degrades any of the desirable properties of monolayered BP, such as the direct band gap at Γ, or the large carrier effective masses. This served as a link to previous theoretical studies researching monolayer BP

which was either encapsulated by hBN [83, 84, 85] or interfaced with other semiconducting materials [86], while our large-scale approach also allowed the simulation of rotated interfaces. Therefore, our large, relatively unstrained, and rotated hBN/BP/hBN heterostructure provided a realistic representation of hBN-encapsulation. We have compared the unfolded spectral function of the independent BP monolayer (Fig. 3.2.a) with that of the BP sheet in hBN/BP/hBN (Fig. 3.2.d). It is clear that no alterations have occurred to the DFT-predicted band gap (0.84 eV), nor to the shape of the bands near the Fermi level. The tendency of DFT to significantly underestimate experimental band gap values [14] was of little consequence to us, since we were only interested in the relative band-gap changes in hBN/BP interfaces. The local density of states (LDOS) plot in Fig. 3.2.d clearly indicates that the hBN/BP/hBP interface has a type-I (straddled) band-alignment, a conclusion which also applies to BP/hBN/BP (Fig. 3.2.e). For both hBN/BP/hBN and BP/hBN/BP (Fig. 3.2.d,e), the black arrows mark band discontinuities; these occur due to interactions between hBN and BP at energies where the hBN-associated DOS is non-zero, i.e. >0.5 eV below the valence band maximum at Γ. Therefore, we find that encapsulation with hBN sheets produce no undesired effects to the electronic structure of pristine monolayer BP. Moreover, it is experimentally known [39] that hBN substrates are usually free of charged impurities and shield from unwanted scattering potentials, while hBN overlayers protect from interactions with air [38, 35, 39]. Consequently, we expect hBN-encapsulation to benefit realistic BP layers, in agreement with the experimentally observed increased carrier mobilities [87], reliable transistor operation in ambient conditions [38], or improved n-type charge transport [35, 39].

Secondly, we have proved that hBN spacers (i.e. BP/hBN/BP interfaces from Fig. 3.2.g) can be utilised to counteract the near-IR to mid-IR absorption shift that occurs in multi-stacked BP sheets[14]. This effect is clearly observed as the transition from a 0.84 eV DFT-predicted band gap in monolayer BP (Fig. 3.2.a) to, for instance, a 0.34 eV gap in the AA-stacked bilayer BP (Fig. 3.2.c). To the best of our knowledge, the electronic-structure implications of such an arrangement have not been studied in any other theoretical work. We started by comparing the spectral function representation of the top BP sheet in the AA-stacked BP bilayer (Fig. 3.2.c) with the spectral function of the top BP sheet in BP/hBN/BP interface (Fig. 3.2.e).

Consequently, in BP/hBN/BP the interaction between the BP sheets is significantly diminished by the monolayer hBN spacer, as evidenced by the almost complete reversion to the monolayer bandstructure of BP. The spacer limits the degeneracy breaking between the states of two BP layers, greatly reducing the splitting of the Γ-point conduction band (CB) valley and valence band (VB) peak in BP/hBN/BP compared to the BP AA-bilayer: 0.14 eV for the CB (0.49 eV for the bilayer) and 0.06 eV for the VB (0.68 for the bilayer). Most importantly, the band gap of the combined system is increased from 0.34 eV in the bilayer case to 0.82 eV in the BP/hBN/BP heterostructure. This is explained by the presence of the hBN spacer, which increases the distance between the BP layers, leading to a stringent reduction in the overlap between their individual states. In essence, a BP bilayer is transformed into a loosely interacting stack of BP single layers. We have thus demonstrated that monolayer BP and BP/hBN/BP interfaces share a similar band gap, which could be used to greatly enhance the external quantum efficiency of BP optoelectronics in the near-IR spectrum, as the use of thicker hBN-spaced BP multilayers instead of BP monolayers would enable increased light absorption. We note that our qualitative conclusions are generally applicable to any BP stacking in the case of hBN-spaced BP multilayers.

Electric fields and TFETs

The residual coupling between the BP layers in BP/hBN/BP, evident from the remaining non-zero splitting of the valence band maximum (VBM) and conduction band minimum (CBM), can be effectively eliminated by using vertical electric fields. In Fig. 3.3 we have plotted the unfolded spectral function and density of states for BP/hBN/BP (with the difference that we have projected on each BP monolayer separately via Eq. 2.3.3), and we have analysed the results as a function of the applied vertical electric field.

In the zero-field state (Fig. 3.3.a), the band edges are split into equally-weighted components in both the bottom and top layers, due to the symmetry of the degeneracy-breaking interactions. However, by applying even a modest vertical field of 0.7 V/nm (Fig. 3.3.b), the spectral weights of the intrusions from one layer into the other decrease dramatically. More specifically, in each split band-pair, the weight of one band decreases by a factor of

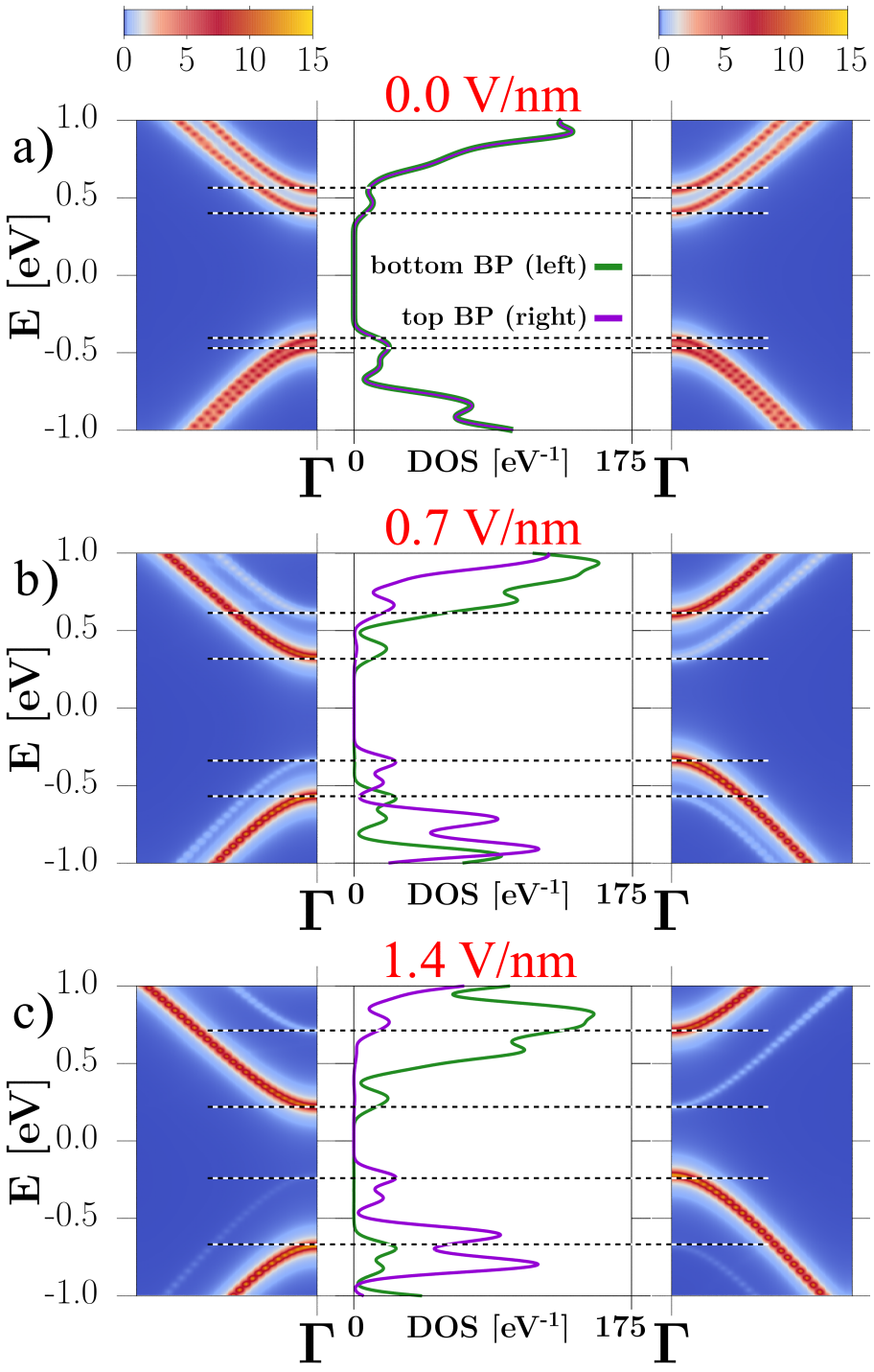


Figure 3.3: Spectral function representations along the Γ-Y direction for the bottom (left insets) and top (right insets) BP layers in BP/hBN/BP, under constant vertical electric fields: 0.0 V/nm (a), 0.7 V/nm (b), and 1.4 V/nm (c); white bands have much lower spectral weight than red bands. Center insets show the projected DOS of each BP layer. Dashed lines highlight the split-band edges.

≈30, while the other doubles in weight, effects which are reversed in the other BP sheet, and which are more pronounced for larger fields (Fig. 3.3.c). In essence, the BP layers are independent in the presence of vertical electric fields. In a realistic setup, substrate effects or impurities would probably break the degeneracy between the BP films even without an external field, achieving the same spectral weight redistribution effect. By vertically displacing the BP films, the hBN spacer induces a larger potential difference between them for a constant field. Thus, the band gap of the composite BP/hBN/BP system is highly tunable under electric fields, showing a linear dependence with a slope of 0.29 eV nm/V; this tunability far exceeds what was previously encountered in pristine BP [73]. The field-induced band shifts in the projected density of states of each layer (Fig. 3.3 - center insets) point toward a spatial separation of charge carriers, with electrons being on the bottom BP layer, while the holes are on the top BP layer.

This careful control of the energy levels achieved through electric fields warranted consideration of the BP/hBN/BP interface as a potential candidate for Tunnelling Field Effect Transistor (TFET) architectures. A model arrangement is shown in Fig. 3.4.a, where the BP/hBN/BP heterostructure is placed between a 4-layer hBN dielectrics and a graphene gates on each side; gate voltages (V_{BG}, V_{TG}) controled the carrier concentrations on each BP layer, while a bias voltage (V_{DS}) was also maintained between the BP films. More details can be found in Appendix B.

In conventional Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs), TMDC channels are generally preferred over BP channels, since the smaller band gap of BP leads to smaller on/off current ratios [12, 16]. A smaller band gap implies that thermal excitations can more easily inject electrons from one layer into the other in the off-state, leading to a non-negligible off-state current and implicitly worse switching capabilities. However, in a TFET arrangement (Fig. 3.4.a), the hBN barrier (Fig. 3.4.b) in between the BP layers strongly inhibits interlayer thermal carrier injection, leaving only quantum tunnelling as a current-generation mechanism. Furthermore, the TFET concept is resurfacing in the context of 2D materials, since layered structures ensure cleaner interfaces, without dislocations or trap states, the main causes for slower switching and increased background current in 3D TFETs. Consequently, our proposal of a BP/hBN/BP TFET

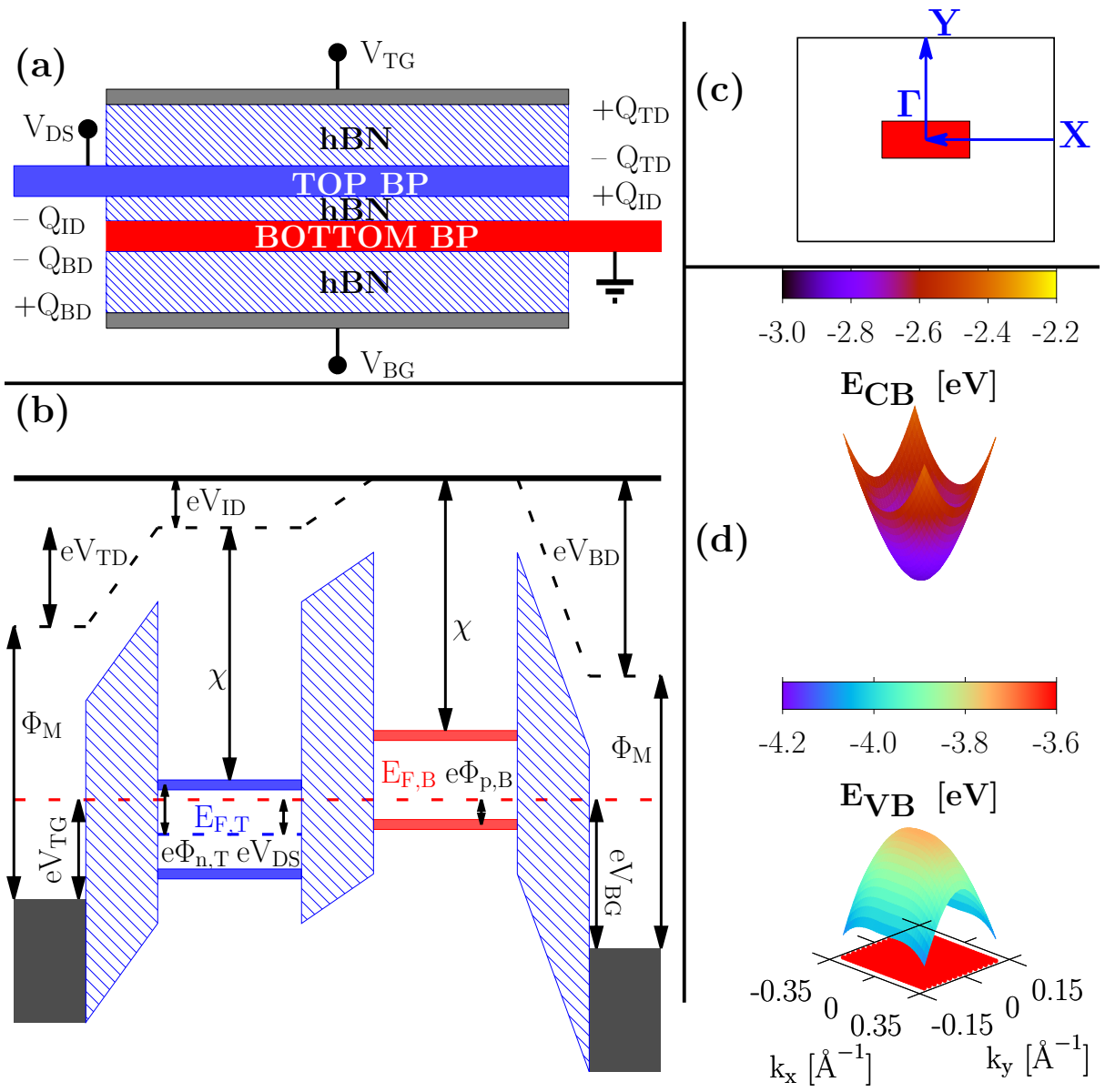


Figure 3.4: (a) Schematic of BP/hBN/BP transistor, also showing induced charges in the dielectrics; corresponding charges in BP have opposite sign. (b) Band diagram, with gate work-function ϕ_M, BP electron affinity χ, and Fermi levels of the top and bottom BP layers $E_{F,T}$ and $E_{F,B}$; $E_{F,B}$ taken as energy reference. (c) Region (red) in the BP primitive-cell Brillouin-zone that was sampled in calculating the energy dispersion relation, as well as the coupling matrix terms. (d) The BP energy levels used in our model (highest valence band and lowest conduction), as seen in the sampled reciprocal-space region.

showed promise due to the superior quality of 2D interfaces, and the possibly reduced thermally-induced currents.

We calculate the tunnelling current in the spirit of the Bardeen formalism [88]:

$$\mathrm{I} = \frac{4\pi e}{\hbar} \sum_{i,j} \sum_{\mathbf{k}_B, \mathbf{k}_T} |\mathrm{O}_{ij}(\mathbf{k}_\mathcal{B}, \mathbf{k}_\mathcal{T})|^2 \, \delta\Big[\mathrm{E}_{i\mathcal{B}}(\mathbf{k}_\mathcal{B}) - \mathrm{E}_{j\mathcal{T}}(\mathbf{k}_\mathcal{T})\Big] \, \Big(\mathrm{f}_{i\mathcal{B}} - \mathrm{f}_{j\mathcal{T}}\Big) \ , \tag{3.1.1}$$

where $\mathbf{k}_\mathcal{B}$ and $\mathbf{k}_\mathcal{T}$ are the in-plane momentum vectors in the bottom ($\mathcal{B}$) and top ($\mathcal{T}$) layers, and $\mathrm{E}_{i\mathcal{B}}(\mathbf{k}_\mathcal{B})$ are eigenvalues of the bottom layer eigenvectors i. The sum over eigenvectors only contains the highest valence band and the lowest conduction band for each individual BP layer (Fig. 3.4.d), and they are defined only in the reciprocal-space region with only one peak (valley) in the valence (conduction) band, i.e. the red shaded area in Fig. 3.4.c. We enforce this constraint through our implicit choice of drain-source and gate voltages, such that most generated electrons and holes, as well as the BP Fermi levels, are at most 0.15 eV below the valence band maximum (VBM) and 0.30 eV above the conduction band minimum (CBM). This boundary is only valid when $|\mathrm{V_{ID}}|$-$\mathrm{E_G}$ (Fig. 3.4) is smaller than the minimum between the aforementioned VB depth and CB height. Moreover, having previously shown that electric fields fully decouple the BP layers in BP/hBN/BP (Fig. 3.4), we can utilise the bandstructure of the independent BP monolayer. Previous works have employed simple parabolic dispersions, but here, with access to the full E-**k** dispersion relation, we can instead sample it explicitly in the relevant region of reciprocal space. The δ-function enforces tunnelling between bands with the same energies, condition which can be relaxed through the use of a Gaussian broadening σ_E instead. The $\mathrm{f}_{i\mathcal{B}}$ and $\mathrm{f}_{j\mathcal{T}}$ terms are bottom and top-layer Fermi occupation factors corresponding to levels i and j (with energies $\mathrm{E}_{i\mathcal{B}}(\mathbf{k}_\mathcal{B})$ and $\mathrm{E}_{j\mathcal{T}}(\mathbf{k}_\mathcal{T})$); their role is to prevent transitions between similarly-occupied levels.

We note that both the shifts in energy and Fermi levels are implicitly dependent on the bias and gate voltages. Their positions are determined by solving a system of non-linear equations for each set of gate and bias voltages, as explained in detail in Appendix B. The coupling matrix term $\mathrm{O}_{ij}(\mathbf{k}_\mathcal{B}, \mathbf{k}_\mathcal{T})$ [88] from Eq. 3.1.1 represents the probability that a carrier can tunnel through a barrier from one momentum state into another:

$$\mathrm{O}_{ij}(\mathbf{k}_\mathcal{B}, \mathbf{k}_\mathcal{T}) = \langle \widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_\mathcal{B}) \mid \mathrm{V} \mid \widetilde{\psi}_{j\mathcal{T}}(\mathbf{k}_\mathcal{T}) \rangle \ , \tag{3.1.2}$$

where V is the scattering barrier potential of hBN and $\widetilde{\psi}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{T}})/\widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})$ are primitive-cell eigenvectors from the top and bottom BP layers, respectively. Similar to the expression for the spectral function representation, Eq. 3.1.2 was evaluated through an unfolding procedure, detailed in Appendix B. One can relax the strict enforcement of in-plane momentum conservation (Eq. 3.1.2) by using a Gaussian broadening of effective width $\sigma_{\mathbf{q}}$, the inverse of which ($\sigma_{\mathbf{q}}^{-1}$) is essentially a real-space coherence length. According to this definition, large coherence lengths are equivalent to small momentum-space Gaussian broadenings, which relate to sharper and larger current peaks upon resonant tunnelling.

Previous studies on TFETs[41, 89, 90] have used simple analytical expressions for the barrier V. In this work, because the full system has been explicitly simulated via LS-DFT, we can utilise a realistic form, obtained from the local potentials (pseudo, Hartree and exchange-correlation). This approach implicitly takes into consideration the interlayer distances and atom corrugation patterns, both of which are accurately described by our chosen ab-initio nonlocal van-der-Waals functional. For these reasons, our coupling matrix terms $\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}})$, derived entirely from DFT-calculated quantities, can be expected to be more realistic than purely analytical expressions.

To make our simulations as relevant as possible to future experiments, we employ rather conservative parameters, indicative of an imperfect system. The presence of random impurities was accounted for by setting the band-energy broadening (σ_E) to 20 meV, room temperature was assumed unless specified, and the coherence length was set to a modest value of $\sigma_{\mathbf{q}}^{-1}$=25.4 nm, well within the range of coherence lengths encountered in other studies[41]. Surprisingly, even in these suboptimal conditions, our simulated BP/hBN/BP TFET shows promise in each of the three studied operating regimes. To ensure a thorough exploration, in Appendix B we have also considered a smaller coherence length ($\sigma_{\mathbf{q}}^{-1}$=6.4 nm) and a larger energy broadening (σ_E = 40 meV).

In the first mode (Fig. 3.5.a), we have utilised our TFET as a Zener diode, in reverse bias. The operating principle is revealed by observing the relative band alignments of the BP films, as shown in the miniature band diagrams from Fig. 3.5; the energy levels that can contribute to tunnelling in the T=0 K limit are highlighted by darker shades of red (for the bottom BP) or blue (for the top BP). In this particular case, the current

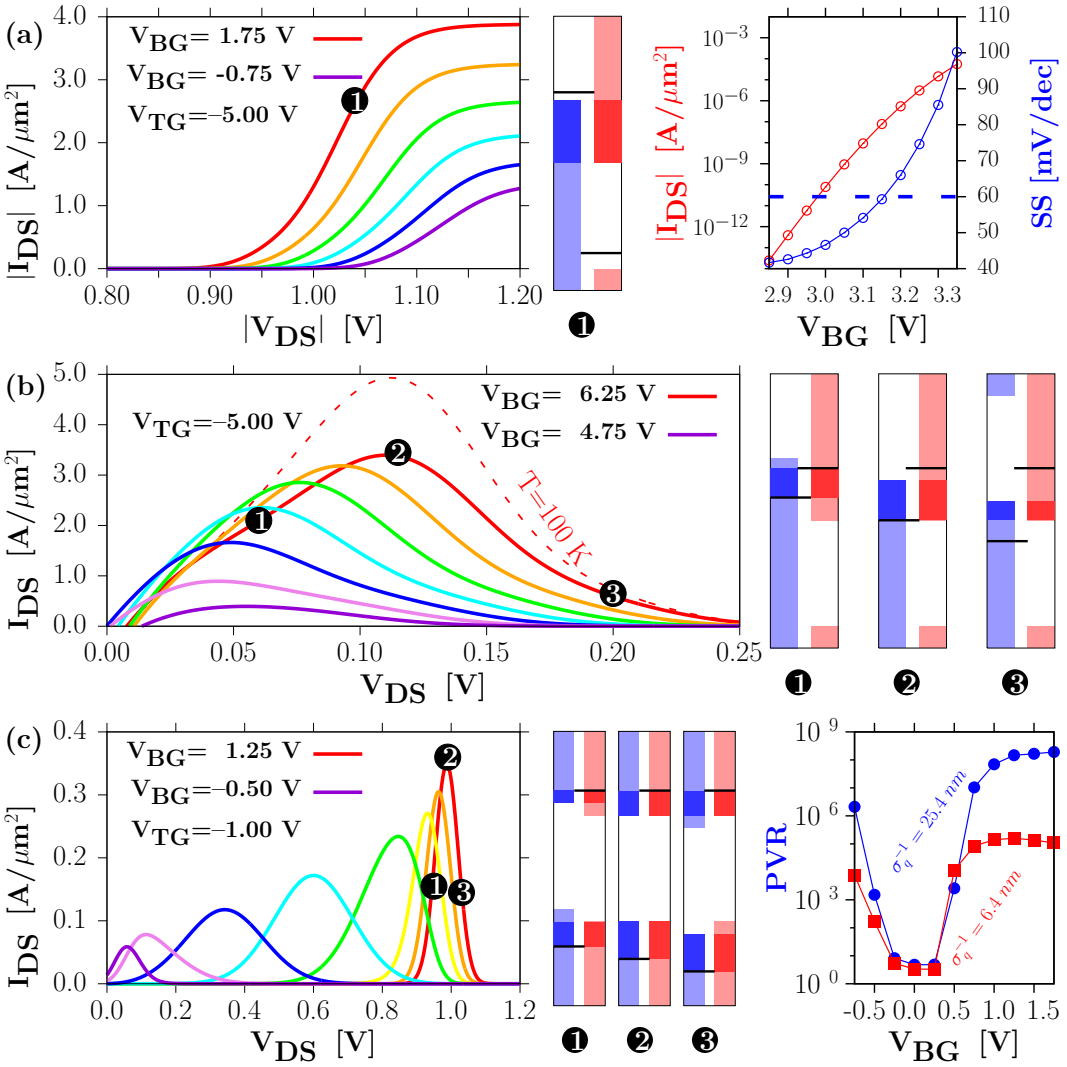


Figure 3.5: Operation modes for the BP/hBN/BP TFET. The top-gate voltage (V_{TG}) is fixed for each case, the back-gate voltage (V_{BG}) is varied in steps of 0.50 V for (a) and 0.25 V for (b),(c). Band diagrams show energy levels of the top (blue) and bottom (red) BP layers – forbidden-energy gaps are white, energy levels involved in tunnelling at 0 K are dark red/blue, Fermi levels in each layer are black lines. (a) Left inset: current dependence on bias voltage for the reverse-bias/Zener regime. Right inset: current (red) and subthreshold swing (blue) dependence on the back-gate voltage, at fixed bias voltage V_{DS}=-0.6 V; dotted blue line shows the theoretical limit for SS in MOSFETs. (b) Left inset: negative differential resistance (NDR) peak for broken-gap arrangement, in forward bias. All solid I-V curves are at 300 K, while the dotted one is at 100 K. (c) Left inset: NDR peak for the aligned-gap arrangement, at low gate voltages. Right inset: peak-to-valley ratio (PVR) for different back-gate voltages, for coherence lengths σ_q^{-1}=25.4 nm and σ_q^{-1}=6.4 nm; the valley current is measured at a bias voltage 0.2 V larger the NDR peak position, while σ_E = 20 meV throughout.

is generated by valence electrons from the p-doped top layer tunnelling to unoccupied conduction states in the p-doped bottom layer. In this regime, the subthreshold swing (SS), defined as the change in gate voltage required to increase the current tenfold, reflects the on/off switching capabilities. In MOSFETs, the thermal carrier injection mechanism limits SS to a minimum value of 60 mV/dec [42]. In contrast, the TFET architecture is not constrained by a lower bound in switching speed, as predicted for TMDC TFETs [89]. In fact, by calculating SS for a range of back gate voltages (right inset of Fig. 3.5.a), we have observed a lower SS than the theoretical limit for MOSFETs, over a current range spanning six orders of magnitude. This essentially implies that a BP/hBN/BP TFET could switch between the on and off state faster than any MOSFET in existence. This shows outstanding potential for high speed electronics, even under our conservative assumptions regarding impurities in BP and lack of in-plane momentum conservation during tunnelling. In our model, the SS is lower than 60 mV/dec only for currents as large as 10^{-5} A/μm^2, when the transistor is not fully on.

The effect of realistic imperfections on this mode (Fig. B.1, Appendix B) could be shown by comparing different combinations of coherence lengths (σ_q^{-1}=25.4 nm and 6.4 nm) and energy broadenings (σ_E=20 meV and 40 meV). We have found that the coherence length has no effect on the general shape of the I-V characteristics or the SS, except for a reduction of the current by a factor of 10. However, a change from σ_E=20 meV to σ_E=40 meV has more significant effects: while the current values are largely maintained, the I-V curves in Fig. B.1 (Appendix B) are much broader, with fewer defined features and peaks. More importantly, the SS is much larger, with its minimum value being 144.4 mV/dec for σ_E = 40 meV, instead of 41.7 mV/dec as for σ_E = 20 meV. Thus, the switching speed is highly correlated to the impurity concentration in BP (linked to σ_E), but rather independent of the BP/hBN interface quality (linked to σ_q^{-1}).

The main characteristic of the second operational mode (Fig. 3.5.b), was the presence of a negative differential resistance (NDR) peak [40]. Such NDR peaks have applications in oscillatory circuits, memory devices, and even multi-valued logic [91]. This mode was revealed by switching to forward bias in the previously discussed regime (Fig. 3.5.a). As before, the band diagrams (Fig. 3.5.b - right insets) indicate a broken gap arrangement,

but now both BP films require degenerate doping for the NDR peak to occur, with the bottom (top) layer being of n (p) type. The tunnelling carriers are clearly the conduction electrons in the bottom layer (dark red shade) and holes in the top layer (dark blue shade). Once the bias is increased, the energy levels of the top layer shift down, until the top-BP Fermi level aligns with the bottom-BP conduction band minimum, leading to a maximum current (point 2 in Fig. 3.5.b). For even higher bias, the forbidden energy gaps truncate the number of tunnelling-capable states, thus decreasing in current (point 3 in Fig. 3.5.b).

Without the hBN interlayer barrier, carriers from the n-side conduction band would eventually be thermally injected into the p-side empty states. This effect would lead to a valley in the current, followed by an increase; this is a behaviour generally observed in Esaki diodes, such as a recent realisation of a 2D BP/$SnSe_2$ diode[92]. Such thermal excitations lead to a lower peak-to-valley ratio in devices exhibiting NDR. However, the hBN barrier shifts this thermally-driven current increase to higher forward-bias voltages, thus extending the range over which the tunnelling current can decrease, and increasing the peak-to-valley ratio. Note that if the temperature is lowered to 100 K, the current peak is more pronounced, due to the sharpening of the Fermi factors in Eq. 3.1.1. Such a behaviour is in perfect agreement with the experimentally-observed temperature dependence of NDR peaks in MoS_2/WSe_2 TFETs [93]. We note that this operation mode exhibits the same dependencies on the coherence length and energy broadening (see Fig. B.2) as described in the previous TFET regime. On the other hand, this mode suffers from low energy efficiency, due to the requirement of large gate voltages. These are needed to enable the degenerate doping of the BP layers and ensure a broken-gap arrangement.

On the other hand, the third operational mode (Fig. 3.5.c), also showing NDR behaviour, has proven to be highly energy efficient, due to the lower required gate voltages. Since the band gaps of the BP films needed to be aligned in this case, there was no need for large bias or gate voltages. While the currents are an order of magnitude lower than in the previous regime, the position of the NDR peak is significantly more tunable under changes in gate voltage, a useful prospect for memory devices. The band diagrams of Fig. 3.5.c

show the two-fold current generation, between the bottom conduction electrons and the top unoccupied conduction states, and from the bottom valence electrons to the top holes. By perfectly aligning the bands through an increase in bias (point 2 in Fig. 3.5.c), a maximum number of tunnelling carriers is obtained, leading to a maximum current. For even larger bias (point 3 in Fig. 3.5.c) the number of carriers is maintained, but the current drops. This is due to the momentum-mismatch caused by the band misalignment, which decreases the coupling matrix terms $|O_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}})|$ between the top ($\mathbf{k}_{\mathcal{T}}$) and bottom ($\mathbf{k}_{\mathcal{B}}$) momentum vectors. Clearly, the change in current due to momentum-mismatch would be most significant in TFETs with same-material films, such as our BP/hBN/BP case or previously studied TMDC/hBN/TMDC architectures [41], and is highly dependent on the coherence length $\sigma_{\mathbf{q}}^{-1}$ (see Appendix B). Since we did not compute thermally-induced currents, we could not precisely pinpoint the value and position of the valley current in the I-V curves from Fig. 3.5.c. However, in the spirit of Campbell *et al.*[41] we have defined the valley current at a fixed offset of 0.2 V from the NDR peak; using this, we could obtain a measure of the peak-to-valley ratio (PVR) dependence on the back-gate voltage.

As shown in the right inset of Fig. 3.5.c, there are two distinct means for large PVR to occur: either at low bias (V_{DS}), for back-gate voltages (V_{BG}) with the same sign as the fixed top-gate voltage (V_{TG}=-1.00 V), or at large bias for back-gate voltages of opposite sign as the top-gate voltage. In the former case, one has the advantage of energy efficiency, due to low required bias; however, the generated peak currents are also low. In the latter case, the peak currents, as well as the PVRs increase dramatically, but at the cost of larger gate bias voltages. Therefore, one has a choice between low-power/low-current and high-power/large-PVR operation, depending on the technological application.

An important observation is that our obtained PVRs are of the same order of magnitude ($\approx 2 \cdot 10^8$) as those predicted by Campbell et al. [41] ($\approx 10^9$) for TMDC/hBN/TMDC TFETs. These results imply that BP can be just as meaningful as TMDCs in TFET applications, as opposed to the case of MOSFETs where BP layers exhibit on/off current ratios [16] three orders of magnitude smaller than TMDC films [94]. Furthermore, BP multilayers have significantly smaller band gaps than TMDCs, requiring only small

voltage changes to switch between the different operational modes presented in our work (Fig. 3.5.a,b,c).

As a last note, a change in coherence length produces profound changes in this operation mode, as shown in Fig. B.3 from Appendix B. Firstly, if the coherence length is reduced fourfold (from 25.4 nm to 6.4 nm), the current decreases by a factor of 500, while the NDR peak current becomes more sensitive to the back-gate voltage. These effects are explained by the enhanced momentum-coupling between alike bands (as in Fig. 3.5.c), as opposed to bands of different character (as in Fig. 3.5.a,b). The peak-to-valley ratios (PVRs) are also highly dependent on the coherence length. For instance, upon a fourfold reduction in coherence length, the maximum PVR is reduced by three orders of magnitude, from $1.92 \cdot 10^8$ to $1.59 \cdot 10^5$. If subsequently the energy broadening is doubled from 20 meV to 40 meV, the maximum PVR is further reduced by another three orders of magnitude, to $1.22 \cdot 10^2$. Therefore, both the current values and the peak-to-valley ratios from the aligned-band NDR regime are highly sensitive to the BP/hBN interface quality and the intrinsic purity of monolayer BP.

3.2 TMDC heterostructures

In this section we discuss the simulations of transition metal dichalcogenide (TMDC) heterostructures, which have been performed in conjunction with our experimental collaborators at the University of Warwick and University of Washington. Most of the results presented here have been published in Ref. [44]. The theoretical results are entirely my contribution, while all experimental data and results belong to my collaborators. In the end, my calculations matched experimental angular-resolved photoemission micro-spectroscopy (ARPES) observations, while also identifying the coexistence of commensurate and incommensurate phases in such interfaces at small twist-angles.

3.2.1 Methodology

The methodology is very similar to what I have previously described in hBN/BP interfaces. Therefore, for individual-layers and aligned-heterostructures, I have once again employed the Quantum Espresso [71] plane-wave DFT package. The monolayer structures were optimised using the ultrasoft potentials of Garrity et al. [70] until forces were smaller than $5{\cdot}10^{-5}$ a.u./bohr ($\approx 2.5{\cdot}10^{-3}$ eV/Å) for monolayers, and $2.5{\cdot}10^{-4}$ a.u./bohr (10^{-2} eV/Å) for bilayers and bulk, while stresses were constrained to be smaller than 0.5 kbar. Odd-numbered TMDC stacks are well known for their strong spin-orbit coupling [95], which induces a large split between energy-bands at the K high-symmetry point (the corner of the hexagonal Brillouin Zone). Subsequently, the bandstructures were calculated using the high-accuracy fully-relativistic PAW potentials of Dal Corso [96], such that spin-orbit interaction was included. As before, we have employed the optB88-vdW [97] functional.

Convergence tests for binding energy as a function of the k-point in-plane sampling (Table 3.6) have revealed the optimum grid to be 9×9 in all TMDCs, with four out-of-plane k-points for the bulk. Moreover, I have utilised a 800 eV plane-wave energy cutoff with a 8000 eV charge density cutoff. The simulation cell height was 30.0 Å, to avoid interaction between periodic images.

As for pinpointing the geometry, I have optimised the lattice parameters of our $MoSe_2$,

Table 3.6: Convergence of binding energy (in meV/bilayer-unit) of bilayer $MoSe_2$, WSe_2, and WS_2 with respect to the in-plane k-point sampling.

in-plane grid	$MoSe_2$	WSe_2	WS_2
6×6	-232.33	-232.54	-222.30
9×9	-231.33	-231.48	-221.45
12×12	-231.36	-231.49	-221.45
15×15	-231.36	-231.49	-221.45

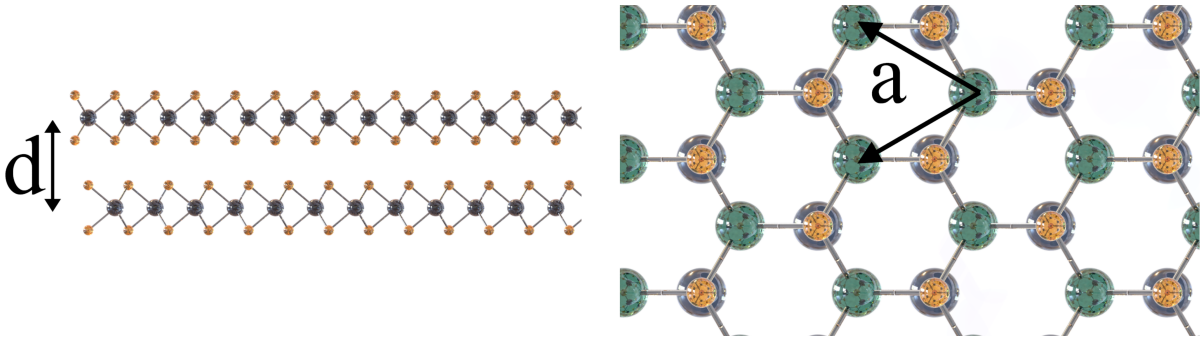


Figure 3.6: Structure of the TMDC layers, side-view (left inset) and top-view (right inset), with lattice vector a and interlayer distance d.

WS_2, and WSe_2 layers, as shown in Table 3.7. TMDC monolayers consists of stacked chalcogen (X) - transition metal (M) - chalcogen (X) films with hexagonal symmetry (space group $P6_3/mmc$), as shown in Fig. 3.6. The primitive unit cell has anti-aligned stoichiometric units (MX_2) in the case of the bulk and bilayer, and a single stoichiometric unit in the case of the monolayer. The distance between the M-planes (d in Fig. 3.6) is generally regarded as the interlayer separation.

Table 3.7: Lattice constants of the $MoSe_2$, WS_2, and WSe_2 monolayers, bilayers and bulk. The bilayers and bulk are in the standard anti-aligned stacking, and d is the distance between the transition metal planes.

		$MoSe_2$	WS_2	WSe_2
monolayer	a [Å]	3.323	3.193	3.324
bilayer	a [Å]	3.326	3.194	3.328
	d [Å]	6.668	6.265	6.555
bulk	a [Å]	3.329	3.196	3.329
	d [Å]	6.593	6.272	6.622

Utilising the well-known experimental lattice constants as reference, there is a good agreement with my obtained values: a 1.2% in-plane and 2.0% out-of plane relative difference

for $MoSe_2$ ($a_{exp} = 3.288$ Å, $d_{exp} = 6.465$ Å)[98], 0.5% in-plane and 0.4% out-of plane difference for WS_2 ($a_{exp} = 3.180$ Å, $d_{exp} = 6.250$ Å)[99], and 1.2% in-plane and 2.1% out-of plane difference for WSe_2 ($a_{exp} = 3.290$ Å, $d_{exp} = 6.485$ Å)[99]. The agreement between the interlayer distances is, as in the case of hBN/BP interfaces (Section 3.1) a testament to the good van-der-Waals description provided by the non-local correlation component of optB88-vdW. Clearly, WS_2/WSe_2 is lattice-mismatched, while $MoSe_2/WSe_2$ has similar lattice parameters, implying that the latter can have a strongly-coupled commensurate phase at 0° rotation.

Similarly to hBN/BP heterostructures, misaligned TMDC interfaces form coincidence cells that are impractical to simulate using plane-wave DFT. Thus, I have once again employed the ONETEP [1] linear-scaling DFT approach. Equivalently to the smaller Quantum Espresso calculations, I have used the same functional (optB88-vdW), kinetic-energy cutoff (800 eV), and PAW potentials [70]. The utilised projector-augmented wave (PAW) datasets of Garrity et al. [70] were non-relativistic, but I have included spin-orbit interaction in a perturbative manner. The advantage of PAW is that one requires a coarser charge-density grid to match the accuracy of ultra-soft pseudopotentials [70]. The Mo and W atoms both contained 14 valence electrons ($4s^2$,$4p^2$6,$4d^5$,$5s^1$ for Mo, $5s^2$,$5p^6$,$5d^4$,$6s^2$ for W), while S and Se contained only 6 valence electrons ($3s^2$,$3p^4$ for S, $4s^2$,$4p^4$ for Se). This resulted in our use of 13 NGWFs for W and Mo (10 for the valence electrons, 3 supporting additional polarisation) and 9 for S and Se (4 for the valence electrons, 5 for additional variational freedom). All NGWFs were chosen to have a large cut-off radius (13.0 bohr), and the convergence criterion was that the root mean square of the NGWF gradient should be $< 2 \cdot 10^{-6}$. For each NGWF conjugate-gradient step, 8 self-consistent density-kernel iterations were performed, without truncation of the density kernel. The geometry optimisation of the heterostructures was performed by relaxing the internal atomic coordinates until the forces were below 0.1 eV/ Å, a reasonable value considering that our systems contained $\approx$ 800 atoms.

In order to slightly reduce the coincidence cells of the overlayed WS_2/WSe_2 and $MoSe_2/WSe_2$ rotated monolayers, I have allowed a 1% strain in WS_2 and $MoSe_2$, a value which was shown not to affect TMDC electronic properties [100]. As previously discussed (Section

3.1), the ONETEP supercell size is equivalent to the k-point sampling. Thus, the convergence test shown in Table 3.6 acts as a requirement for each layer in our interfaces to contain at least 81 unit cells (243 atoms).

3.2.2 WS_2/WSe_2 and $MoSe_2/WSe_2$

I started by observing the band-structure changes upon stacking different TMDC layers, in particular the WS_2/WSe_2 and $MoSe_2/WSe_2$ interfaces, the structures of which are shown in Figs. 3.7.a,d. The supercell was constructed such that the rotation angle is non-zero but small, matching the experimentally induced twist, while trying to also limit the structure size. Taking into account the previously discussed constraints, the supercells have 762 (363 for WSe_2, 399 for WS_2) and 873 atoms (432 for WSe_2, 441 for $MoSe_2$), with the layers being rotated by 4.3° (WS_2/WSe_2) and 8.2° ($MoSe_2/WSe_2$), respectively..

There is no need to pricisely match the experimental angles, as in a previous publication I have shown that the interlayer coupling is almost invariant to rotation in such lattice-mismatched heterostructures [36]. In order to observe the bandstructure effects of each layer in the presence of the other, I have calculated the unfolded spectral function (as described in Section 2.3), which was projected selectively on each of the component layers, as shown in Fig. 3.7.b,c,e,f. I have assumed that the independent and stacked versions of each monolayer have the same energy at their respective valence band maximum (VBM) at K - this shall also be proven later in the section. My reasoning is that bands at K bands are known to be rather independent of interlayer stacking, since they are generated by intra-plane interactions between the transition metal d-orbitals and chalcogen p_x and p_y functions.

Firstly, Figs. 3.7.c,f reveal that all independent monolayers have a direct band gap at their respective K-points, as is already well known [44]. The energy splitting at K induced by spin-orbit coupling is replicated in my calculations, with the values (0.18 eV and 0.47 eV for monolayered $MoSe_2$ and WSe_2) being in agreement with experiment [44] (0.24 eV and 0.49 eV, respectively). This shows that the perturbative spin-orbit coupling implementation in ONETEP is well behaved.

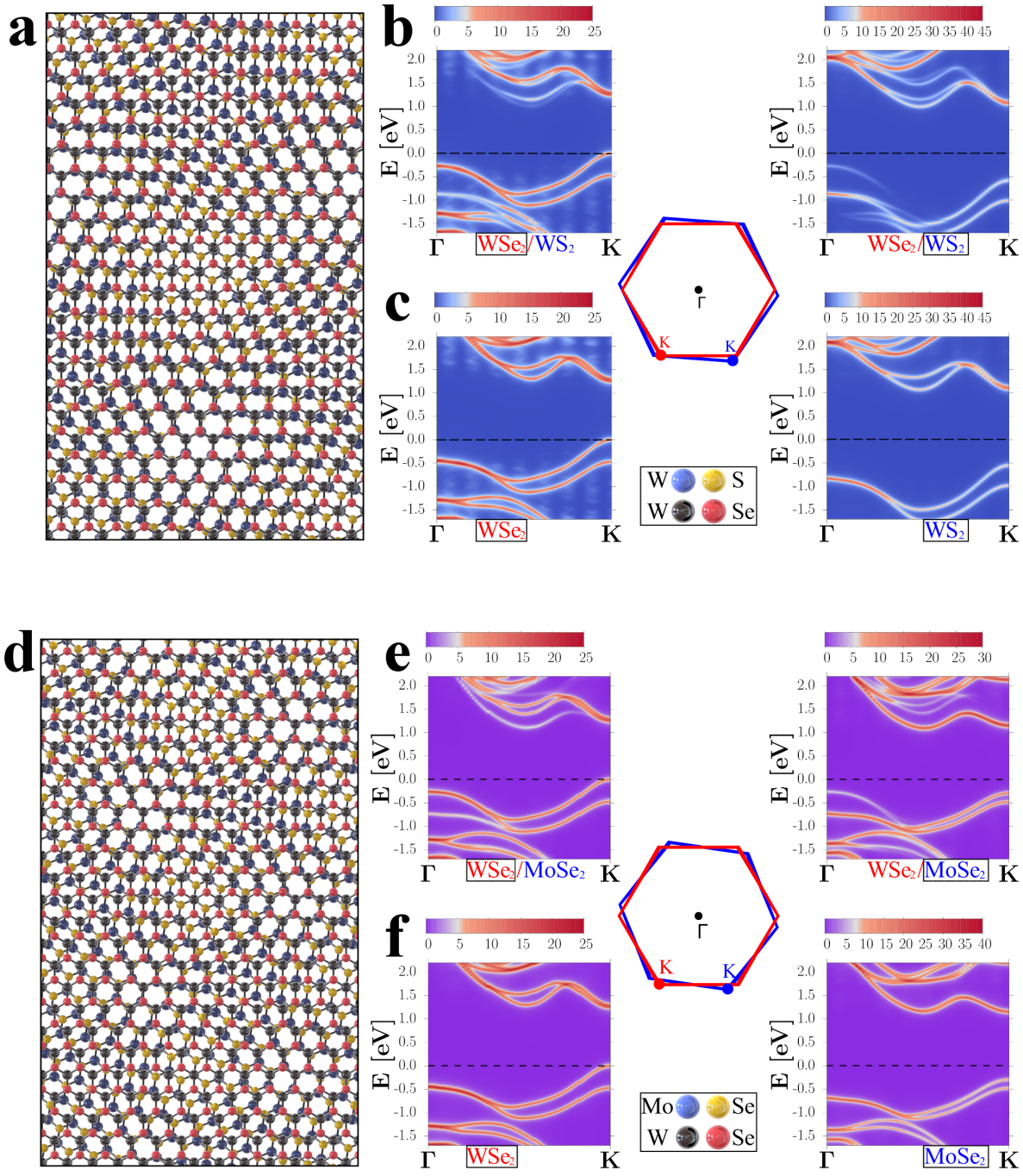


Figure 3.7: Structure of the misaligned WSe_2/WS_2 (a) and $MoSe_2/WSe_2$ interfaces (d). Unfolded spectral function projected on the component WSe_2 (b-left) and WS_2 (b-right) monolayers, as well as on the independent WSe_2 (c-left) and WS_2 (c-right) monolayers. Similarly for the WSe_2 (e-left) and $MoSe_2$ (e-right) monolayers and their stand-alone counterparts (f). The Brillouin zones are rotated by 4.3° and 8.2°, respectively. The energy reference is the energy of the valence band maximum (VBM) at K in the heterostructure.

By comparing Fig. 3.7.b with c, and Fig. 3.7.e with f, it is clear that when independent films are stacked, there are low spectral-weight bands (the white ones) protruding from one monolayer into the other. Moreover, Figs. 3.7.b and Figs. 3.7.c, reveal that stacking interactions raise the valence band maximum (VBM) at Γ of WSe_2 by 200 meV, and lower the VBM of WS_2 by 20 meV. This trend is in agreement with the experimental findings of our collaborators, while the values were found to be 150 meV and 50 meV, respectively. Similarly, Figs. 3.7.e,f show that the VBM at Γ of WSe_2 is raised by 202 meV, while the $MoSe_2$ VBM is lowered by 67 meV, close to the experimental 250 meV and 90 meV, respectively [44].

In both interfaces, the global valence band maximum (VBM) is at the K of WSe_2, while the global conduction band minimuma (CBM) are at the K of WS_2 and the K of $MoSe_2$, respectively - this implies that both WS_2/WSe_2 and $MoSe_2/WSe_2$ are type II heterostructures. Since the Γ-K directions are rotated, the positions of the VBM and CBM form an indirect gap; however, when compared to homo-structural TMDC multilayers, which generally have a Γ - Λ (halfway point between Γ and K) indirect band-gap, the momentum-space separation between the CBM and VBM is much smaller in the heterostructured interface. Note that the valence band offset at K between WSe_2 and WS_2 is 0.55 eV, in good agreement with the 0.62 eV obtained by Kang et al. (using the hybrid HSE functional [101]). However, the valence band offset between WSe_2 and $MoSe_2$ is 0.29 eV, significantly lower than the 0.43 eV predicted by Kang et al. [101], but in perfect agreement with the experimentally observed 0.30 eV [44]. Thus, interlayer interactions induced by stacking are crucial for the correct modelling of band offsets, with the independent-layer approximation proving to be insufficient.

3.2.3 Commensurate and incommensurate domains

Experimentally, in the $MoSe_2/WSe_2$ interface with a low twist angle ($\approx 1°$) one observes an unusual behaviour: the presence of three highest valence bands close in energy at Γ (Fig. 3.8), instead of the expected two spin-degenerate bands for the bilayer mono-$MoSe_2$/mono-WSe_2 [44], and four bands instead of the three spin-degenerate levels for the trilayer bi-$MoSe_2$/mono-WSe_2. My theory is that commensurate/incommensurate

transitions occur in the micro-flake, as previously observed for graphene/hBN [102]. Since the flake orientation is very close to 0°, there would be regions where $MoSe_2$ and WSe_2, having almost identical lattice parameters, are perfectly commensurate and aligned, and regions where they are misaligned, and thus significantly less coupled.

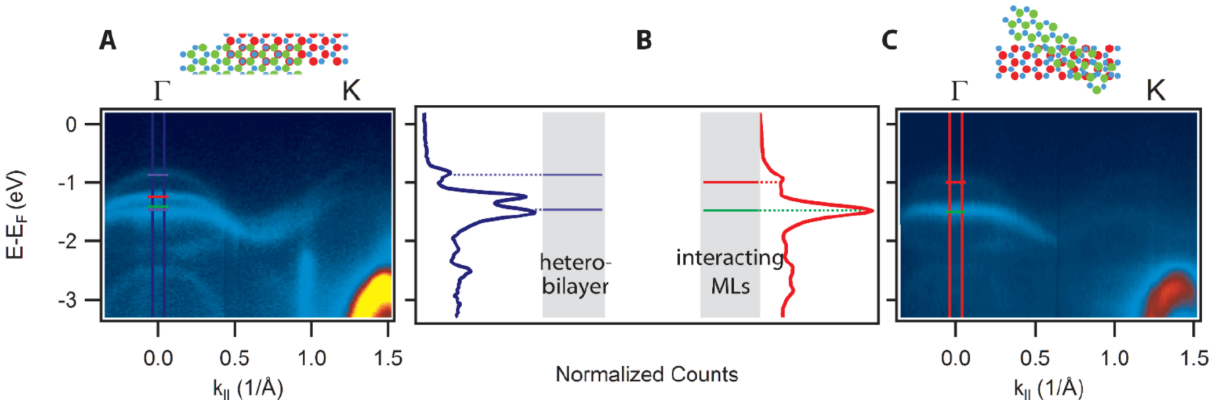


Figure 3.8: Courtesy of Wilson et al. [44]: ARPES slice along Γ-K for a nearly-aligned (A) and misaligned (C) $MoSe_2/WSe_2$ interface. Energy distribution curve at Γ (B). The DFT predicted band positions at Γ of the purely commensurate heterobilayer (two bands, marked in blue), and interacting monolayers (two bands, marked in red for WSe_2 and green for $MoSe_2$) are also shown in (B), along with the band positions at Γ for the isolated monolayers (two bands, marked in red for WSe_2 and green for $MoSe_2$).

As I have previously shown [36], if one has a bilayer structure formed out of materials with the same lattice constant, there is significant coherence between the layers only in the perfectly aligned or anti-aligned cases (including any lateral translations between the layers). If the layers are rotated even slightly from perfect alignment, they become decoupled to a large extent, but the average electronic densities still interact. The latter case would imply significantly less repelling (i.e. smaller energy difference) between the highest valence bands at Γ, as compared to the perfectly commensurate case.

The commensurate phase can easily be simulated in Quantum Espresso, since the monolayer $MoSe_2$ (a = 3.323 Å), bilayer $MoSe_2$ (a = 3.326 Å) and WSe_2 (a = 3.324 Å) have almost identical same in-plane lattice constants (Table 3.7). Thus, the bandstructures of the anti-aligned mono-$MoSe_2$/mono-WSe_2 and bi-$MoSe_2$/mono-WSe_2 interfaces are shown in Fig. 3.9.

Clearly, in the commensurate case one observes the intuitive number of spin-degenerate highest valence bands at Γ: two for mono-$MoSe_2$/mono-WSe_2 and three for bi-$MoSe_2$/mono-

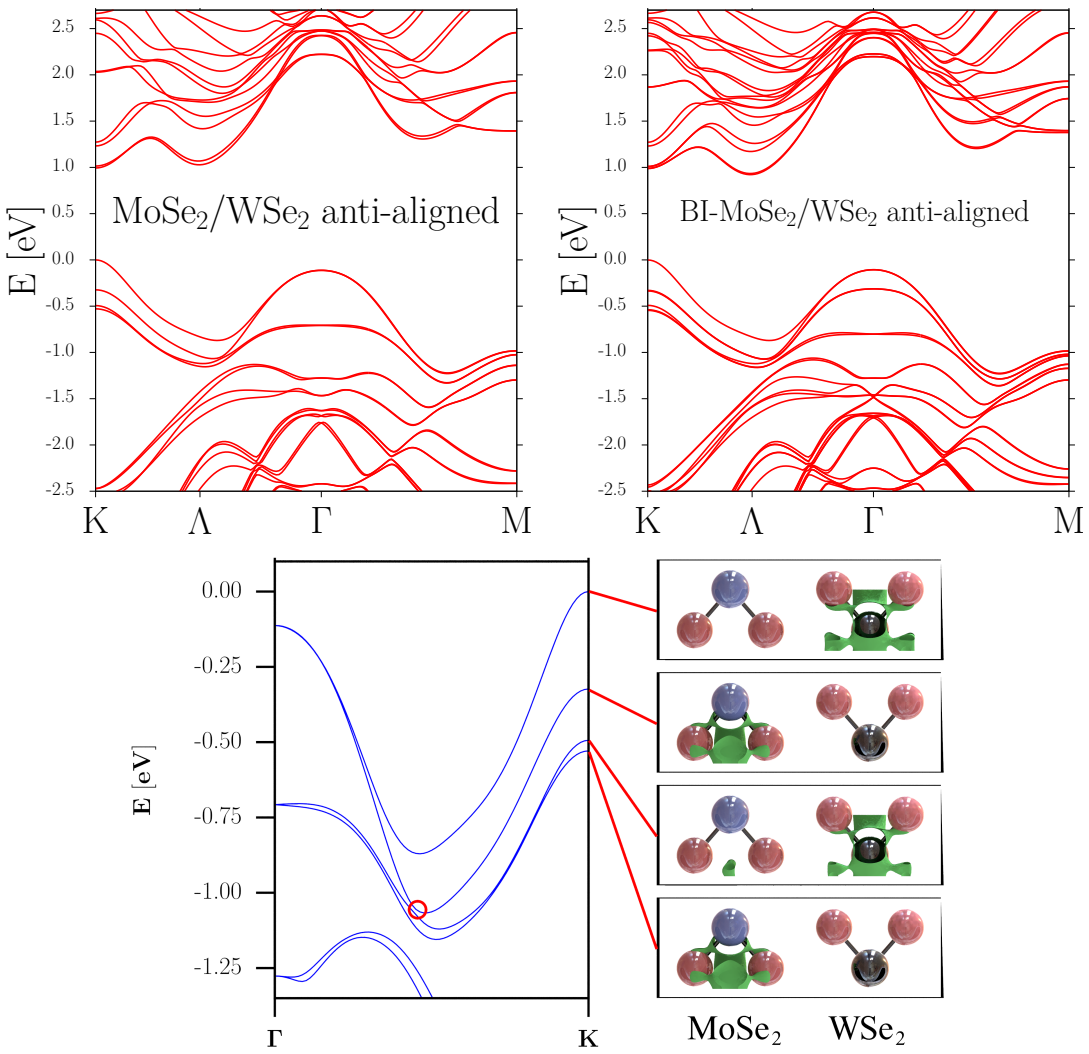


Figure 3.9: Bandstructure of the perfectly anti-aligned mono-$MoSe_2$/mono-WSe_2 (top-left) and bi-$MoSe_2$/mono-WSe_2 (top-right). Charge density isosurfaces (bottom) at the four highest valence bands at for K point in mono-$MoSe_2$/mono-WSe_2. The crossing occurs at the encircled red point, but it is not visible in the bandstructure due to the plotting method.

WSe_2. In both cases the VBM is at K; however, while the CBM for mono-$MoSe_2$/mono-WSe_2 is at K, for bi-$MoSe_2$/mono-WSe_2 it is at Λ (65 meV lower than the CBM at K), making the latter an indirect band-gap heterostructure. As a last note on the commensurate-case bandstructures, one observes that at the K point of mono-$MoSe_2$/mono-WSe_2, the bands corresponding to each layer are crossed, as shown by the isosurfaces in Fig. 3.9.c. The aforementioned isosurfaces clearly show that the spin-split highest VBs at K are generated by the hybridisation of the W or Mo d-orbitals and the S or Se p_x and p_y functions, not by interactions between the layers. We observe only a slight increase in

the splitting of each spin-polarised band corresponding to each layer: 205 meV for $MoSe_2$ (compared to 188 meV in the independent layer) and 494 meV in WSe_2 (compared to 479 meV in the independent layer). These values illustrate the reduced effect that stacking has on the spin-orbit induced splitting at K - this is a step towards proving the assumption that the VBM at K of monolayers can be used as reference even in stacked forms.

For the misaligned case, I start with a simple but instructive approximation. I have shown in the WS_2/WSe_2 interface (Fig. 3.7) and other situations [36] that in spite of the reduced interlayer coupling in incommensurate heterostructures, hybridisation between bands of the stacked monolayers is always present near Γ. However, one could naively approximate the bandstructure of misaligned stacks as a simple superposition of the independent layers. This can easily be simulated in Quantum Espresso by using the commensurate heterostructure cases, and displacing the component layers further apart until the bandstructure stops changing. The results of this procedure are shown in Fig. 3.10. In the end, we can apply the necessary corrections to the independent-layer band edge positions, as obtained from the ONETEP calculations on the misaligned $MoSe_2/WSe_2$ interface (Fig. 3.7).

As previously discussed, the energy reference for the spectral function unfolding calculations (Fig. 3.7) was taken to be the VBM at K. Indeed, upon the transition from the maximally coupled ("MC") state ($\Delta d = 0$ Å in Fig. 3.10) to the decoupled ("D") state ($\Delta d >= 2.0$ Å), the reference value varies by less than 40 meV. I also consider the ONETEP corrections applied to the fully decoupled state, which will be denoted as "D+". As a reminder, these corrections specify that in mono-$MoSe_2$/mono-WSe_2, the E^{Γ}_{HOMO} of WSe_2 is up-shifted by 202 meV in misaligned stacks, while E^{Γ}_{HOMO} of $MoSe_2$ (i.e. $E^{\Gamma}_{\text{HOMO-1}}$ of the heterostructure) is down-shifted by 67 meV.

Starting with mono-$MoSe_2$/mono-WSe_2 (Fig. 3.10-top), it is clear that in both the "D" and "D+" cases, the energy differences between VBMs of the monolayers at Γ are significantly smaller than the ones at the "MC" state. The second-highest VB at Γ ($E^{\Gamma}_{\text{HOMO-1}}$) remains at roughly the same position in the "MC", "D" and "D+" cases, while the highest VB at Γ (E^{Γ}_{HOMO}) is by 340 meV lower in "D" or 138 meV lower in "D+" than it was at the optimum interlayer distance. Thus, considering a superposition between the "MC"

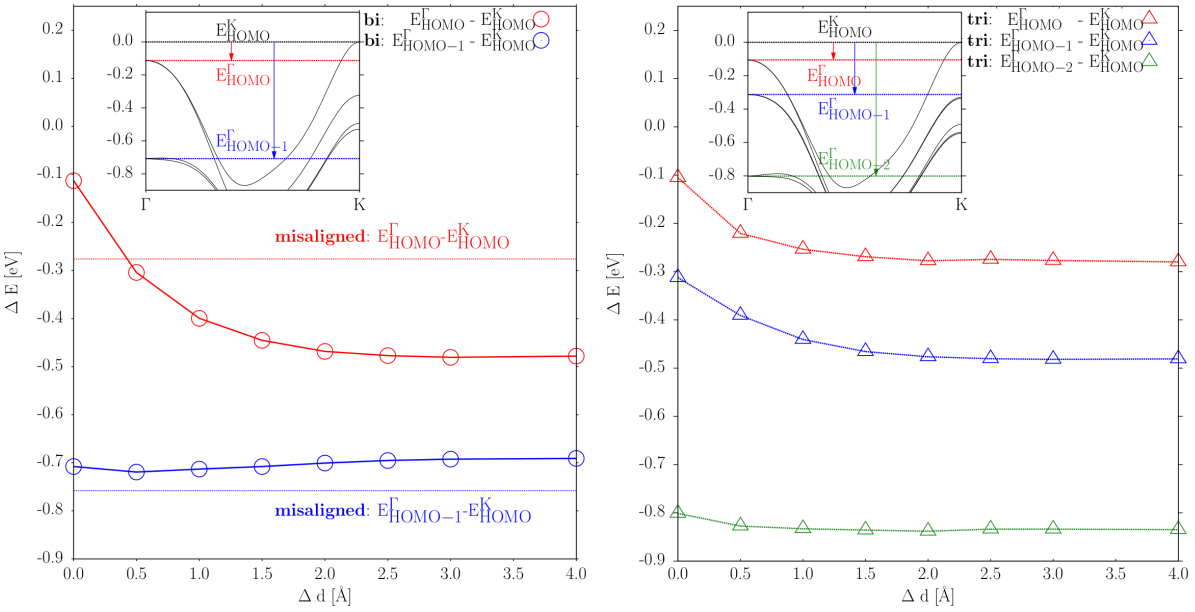


Figure 3.10: Changes in the highest VB energy at Γ for the perfectly anti-aligned mono-$MoSe_2$/mono-WSe_2 bilayer (left) and bi-$MoSe_2$/mono-WSe_2 trilayer (right), as the layers are pulled apart. The energy reference is the highest VB (HOMO) at K for each case, the absolute energy of which changes by < 40 meV under displacement. Each plot contains an inset showing the valence bands and energy levels of interest for the commensurate mono-$MoSe_2$/mono-WSe_2 (left) and bi-$MoSe_2$/mono-WSe_2 (right). The dashed lines (denoted by "**misaligned**" in the top plot represent the decoupled levels after being corrected by the ONETEP calculations on rotated mono-$MoSe_2$/mono-WSe_2.

and "D" or "D+" cases, one would indeed observe 3 bands in the ARPES measurements: the "D"/"D+"-state E^{Γ}_{HOMO}, the MC-state "E^{Γ}_{HOMO}", and lastly, a nearly-degenerate combination of $E^{\Gamma}_{\text{HOMO-1}}$ from the "MC" and "D"/"D+" situations. Consequently, for mono-$MoSe_2$/mono-WSe_2, my prediction of a commensurate/incommensurate superposition is in perfect agreement with the experimental ARPES results.

In the case of the bi-$MoSe_2$/mono-WSe_2 (Fig. 3.10-bottom), however, the situation is slightly different. $E^{\Gamma}_{\text{HOMO-2}}$ maintains a relatively constant position between the "MC" and "D" cases. On the other hand, as the layers interact less, E^{Γ}_{HOMO} and $E^{\Gamma}_{\text{HOMO-1}}$ both decrease by 140 meV while maintaining a constant energy difference of 210 meV between them. Therefore, at first glance, by superimposing the HOMO, HOMO-1, and HOMO-2 from the "MC" and "D" cases, one would obtain five spin-degenerate bands, instead of the experimentally observed four: E^{Γ}_{HOMO} of "MC", E^{Γ}_{HOMO} of "D", $E^{\Gamma}_{\text{HOMO-1}}$ of "MC", $E^{\Gamma}_{\text{HOMO-1}}$ of "D", and the nearly-degenerate bands of the "MC"-state and "D"-state

$E^{\Gamma}_{\text{HOMO-2}}$. However, the "D"-state E^{Γ}_{HOMO} and "MC"-state $E^{\Gamma}_{\text{HOMO-1}}$ are only separated by 60 meV. This small energy difference, accompanied by potential corrections from the realistic ONETEP calculations on this trilayer, might result in the near-degeneracy of the "D" E^{Γ}_{HOMO} and "MC" $E^{\Gamma}_{\text{HOMO-1}}$. This would explain the experimentally-observed four bands instead of the intuitive five, further strengthening our prediction of a commensurate/incommensurate transition occuring in $MoSe_2/WSe_2$ heterostructures. No ONETEP calculations were performed for the misaligned trilayer, due to their large cost on this large structure given the employed methodology. However, the ONETEP corrections did not change the qualitative picture for mono-$MoSe_2$/mono-WSe_2, so one would expect that the "MC" and "D" states for bi-$MoSe_2$/mono-WSe_2 are sufficient to describe the underlying physical principle.

In conclusion, I have shown that such commensurate and incommensurate domains can occur in $MoSe_2/WSe_2$ heterostructures with low but non-zero twist angles. This enables access to different electronic-structures within the the same flake, surely an interesting concept for integrated circuit applications.

Chapter 4

Linear-response phonons

4.1 Introduction

As discussed in the introduction, one of the long term goals of the ONETEP community is to implement linear-scaling Raman spectroscopy capabilities. As a brief overview, in the case of nonresonant Raman scattering, the Stokes process involves an incoming photon (frequency ω_I, polarisation $\mathbf{e}_I$) being scattered to an outgoing photon (frequency $\omega_O = \omega_I - \omega_P$, polarisation $\mathbf{e}_O$) by creating a phonon of frequency ω_P. Thus, in order to obtain an actual Raman spectrum, one needs to know the Raman tensor α_P corresponding to the previously mentioned induced phonon mode with frequency ω_P (details in Ref. [103]):

$$\alpha_P^{ij} \propto \sum_{k',\beta} \frac{d^3E}{d\tau_{k'\beta}\, d\epsilon_i\, d\epsilon_j} \epsilon_i \epsilon_j u_P^{k'\beta} \ , \tag{4.1.1}$$

where E is the total energy of the system (including any external potentials), $\frac{d}{d\tau_{k'\beta}}$ is a perturbation of the atomic position of atom k' in the direction β, $\frac{d}{d\epsilon_i}$ and $\frac{d}{d\epsilon_j}$ are perturbations due to external electric fields, and $u_P^{k'\beta}$ is the eigenvector of the phonon created by perturbing atom k' along β.

It is thus clear that for the Raman tensor we first need the third-order change in energy (with respect to the perturbation of two electric fields and one atomic displacement), and also the phonon eigenvectors. Since our goal is to make the computation of Eq. 4.1.1 as fast as possible, we need to implement a fast and efficient method for computing the phonon eigenvectors and the third-order energy terms. In this thesis, as a starting step to the implementation of Raman spectroscopy, we concentrate on obtaining the phonon frequencies and eigenvectors in the LS-DFT framework of the ONETEP code [1].

While ONETEP already has a finite-differencing (FD) phonon scheme, this method is not suitable for use in large systems, such as our 2D heterostructures. Therefore, we chose

to compute phonon frequencies in the density-functional perturbation theory (DFPT / linear-response) framework. At zero phonon wavevectors, a linear-response calculation has the same runtime as a finite-differencing scheme, due to the identical number of perturbations. However, density functional perturbation theory is significantly faster at non-zero wavevectors, since one does not require supercells to obtain dynamical matrices at arbitrary phonon wavevectors. We will develop this linear-response scheme for the projector augmented wave (PAW) formalism, in insulators, in the LS-DFT framework of ONETEP.

The aforementioned phonon mode frequencies (ω_P) and eigenvectors ($u_P^{k'\beta}$ in Eq. 4.1.1) can be obtained from the generalised eigenvalue problem:

$$\sum_{k',\beta} D^{k\alpha,k'\beta}(\mathbf{q}) u_P^{k'\beta}(\mathbf{q}) = \omega_P^2 u_P^{k\alpha}(\mathbf{q}) \ , \tag{4.1.2}$$

where $\mathbf{q}$ is the phonon wave-vector, k and k' are atomic indices and α, β are the perturbation directions, while $D^{k\alpha,k'\beta}(\mathbf{q})$ are terms of the dynamical matrix, expressed in general as:

$$D^{k\alpha,k'\beta}(\mathbf{q}) = \frac{1}{\sqrt{M_k M_{k'}}} \frac{1}{N_{\text{cells}}} \sum_{\mathbf{R},\mathbf{R}'} e^{-i\mathbf{q}\cdot\mathbf{R}} \Phi_{k\alpha\mathbf{R},k'\beta\mathbf{R}'} \ e^{i\mathbf{q}\cdot\mathbf{R}'} \ , \tag{4.1.3}$$

where $\mathbf{R}, \mathbf{R}'$ are multiples of the simulation cell lattice vectors, N_{cells} is the number of considered cells, and $\Phi_{k\alpha\mathbf{R},k'\beta\mathbf{R}'}$ are force-constant matrix elements. The latter term is the second-order energy changes due to atomic displacements performed in cells with origins at $\mathbf{R}$ and $\mathbf{R}'$, respectively:

$$\Phi_{k\alpha\mathbf{R},k'\beta\mathbf{R}'} = \frac{d^2 E}{d\tau_{k\alpha\mathbf{R}} \ d\tau_{k'\beta\mathbf{R}'}} \tag{4.1.4}$$

Therefore, in the following we will derive the force constant matrix terms in ONETEP, within the PAW framework, for insulators. Also, since ONETEP is utilised for simulating large systems, the considered ground-state wavefunctions have a $\mathbf{0}$ wave-vector (Γ-point), unless specified otherwise.

Lastly, we note that our implementation differs from DFPT implementations in other plane-wave codes (VASP [104], QuantumEspresso [105, 106, 107] or Abinit [108]), since we avoid the use of eigenstates in order to preserve linear scaling. As the reader will see, this approach requires novel optimisation methods and a reformulation of the theoretical

framework. Firstly, in subsections 4.2.1 – 4.2.3 we build upon the Abinit [108] implementation with the addition that we also include q-wavevector dependencies. The rest of the chapter is exclusively our novel work, in which we adapt the DFPT formalism to the ONETEP framework of density kernels and adaptive local orbitals.

4.2 Force-constant matrix

4.2.1 Preliminaries

To determine the second order energy, one requires the first order wavefunction ([108]), as variation-perturbation theory dictates that the n-th order wavefunction is sufficient for the knowledge of the energy of order $2n$ and $2n+1$. Therefore, for $\mathbf{q}=\mathbf{0}$ we expand the pseudo (PS) wavefunction as:

$$|\widetilde{\psi}_i[\lambda,\epsilon]\rangle = |\widetilde{\psi}_i\rangle + \sum_{\mathbf{R}} \lambda_{\mathbf{R}} |\widetilde{\psi}_i^{1\lambda_{\mathbf{R}}}\rangle + \sum_{\mathbf{R}'} \epsilon_{\mathbf{R}'} |\widetilde{\psi}_i^{1\epsilon_{\mathbf{R}'}}\rangle \ , \tag{4.2.1}$$

where $\lambda_{\mathbf{R}}, \epsilon_{\mathbf{R}'}$ are each perturbations (of one or multiple atoms, each in a certain direction) performed only in cell-images displaced by $\mathbf{R}$ and $\mathbf{R}'$. The perturbation-dependencies of the wavefunctions are indicated by "[ ]", and the index i takes values in the range $[0 : N_e/2]$, due to spin-degeneracy. A perturbation of the system (for instance $\epsilon_{\mathbf{R}'}$) is equivalent to applying $\hat{u}_\epsilon \cdot \frac{d}{d\mathbf{R}_\epsilon}$ in a cell with the origin at $\mathbf{R}'$, where $\hat{u}_\epsilon$ is the perturbation direction unit vector, and $\mathbf{R}_\epsilon$ is the position of the moved atom within cell $\mathbf{R}'$. This means that ϵ and λ are combined indices, denoting both the atom and the direction of the perturbation. While $\frac{d}{d\lambda_{\mathbf{R}}}$ and $\frac{d}{d\epsilon_{\mathbf{R}'}}$ are real displacements, the $\mathbf{q}$-vector counterparts

$$\frac{d}{d\lambda^*_{\mathbf{q}}} = \sum_{\mathbf{R}} e^{-i\mathbf{q}\cdot\mathbf{R}} \frac{d}{d\lambda_{\mathbf{R}}} \quad \text{and} \quad \frac{d}{d\epsilon_{\mathbf{q}}} = \sum_{\mathbf{R}'} e^{i\mathbf{q}\cdot\mathbf{R}'} \frac{d}{d\epsilon_{\mathbf{R}'}} \tag{4.2.2}$$

are complex quantities obeying the equivalence $\lambda^*_{\mathbf{q}} = \lambda_{-\mathbf{q}}$ and $\epsilon^*_{\mathbf{q}} = \epsilon_{-\mathbf{q}}$.

One can easily extend the wavefunction expansion from Eq. 4.2.1 to non-zero q-points:

$$\begin{aligned}
|\widetilde{\psi}_i[\lambda_{\mathbf{q}},\epsilon_{\mathbf{q}}]\rangle &= |\widetilde{\psi}_i\rangle + \lambda_{\mathbf{q}} |\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}\rangle + \lambda_{-\mathbf{q}} |\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}\rangle + \epsilon_{\mathbf{q}} |\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \epsilon_{-\mathbf{q}} |\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}\rangle \\
\langle\widetilde{\psi}_i[\lambda_{\mathbf{q}},\epsilon_{\mathbf{q}}]| &= \langle\widetilde{\psi}_i| + \lambda^*_{\mathbf{q}} \langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}| + \lambda^*_{-\mathbf{q}} \langle\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}| + \epsilon^*_{\mathbf{q}} \langle\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}| + \epsilon^*_{-\mathbf{q}} \langle\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}| \\
&= \langle\widetilde{\psi}_i| + \lambda_{-\mathbf{q}} \langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}| + \lambda_{\mathbf{q}} \langle\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}| + \epsilon_{-\mathbf{q}} \langle\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}| + \epsilon_{\mathbf{q}} \langle\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}|
\end{aligned} \tag{4.2.3}$$

Therefore, the first-order wavefunctions themselves are defined as:

$$|\psi_i^{1\epsilon_{\mathbf{q}}}\rangle = \frac{d}{d\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle = \frac{1}{\sqrt{N_{\text{cells}}}}\sum_{\mathbf{R}'} e^{i\mathbf{q}\cdot\mathbf{R}'}\,|\widetilde{\psi}_i^{1\epsilon_{\mathbf{R}'}}\rangle \quad \text{and}$$
$$\langle\psi_i^{1\epsilon_{\mathbf{q}}}| = |\psi_i^{1\epsilon_{\mathbf{q}}}\rangle^\dagger = \frac{d}{d\epsilon_{-\mathbf{q}}}\langle\widetilde{\psi}_i| = \frac{1}{\sqrt{N_{\text{cells}}}}\sum_{\mathbf{R}'} e^{-i\mathbf{q}\cdot\mathbf{R}'}\,\langle\widetilde{\psi}_i^{1\epsilon_{\mathbf{R}'}}| \tag{4.2.4}$$

Note that the conjugate transpose of $\langle\mathbf{r}|\psi_i^{1\epsilon_{\mathbf{q}}}\rangle$ is indeed denoted as $\langle\psi_i^{1\epsilon_{\mathbf{q}}}|\mathbf{r}\rangle$, while

$$\langle\mathbf{r}|\psi_i^{1\epsilon_{\mathbf{q}}}\rangle = \langle\mathbf{r}|\psi_i^{1\epsilon_{-\mathbf{q}}}\rangle^\dagger = \langle\psi_i^{1\epsilon_{-\mathbf{q}}}|\mathbf{r}\rangle \tag{4.2.5}$$

due to the fact that the ground-state wavefunction $|\widetilde{\psi}_i\rangle$ is at Γ, meaning that $\mathbf{q}$ is the only wave-vector present.

In general, it is crucial in the context of density-functional perturbation theory to work with phase-factorised quantities, which have the same periodicity as the underlying lattice. To obtain this, we define phase-factorised position-dependent quantities (with a bar overhead) as

$$\langle\mathbf{r}|\overline{\psi_i^{1\epsilon_{\mathbf{q}}}}\rangle = e^{-i\mathbf{q}\cdot\mathbf{r}}\,\langle\mathbf{r}|\psi_i^{1\epsilon_{\mathbf{q}}}\rangle \quad \text{and} \quad \langle\overline{\psi_i^{1\epsilon_{\mathbf{q}}}}|\mathbf{r}\rangle = e^{i\mathbf{q}\cdot\mathbf{r}}\,\langle\psi_i^{1\epsilon_{\mathbf{q}}}|\mathbf{r}\rangle \tag{4.2.6}$$

One could repeat the perturbative expansion from Eq. 4.2.1 in the densities, starting with the pseudo-density:

$$\begin{aligned}\widetilde{\rho}[\lambda_{\mathbf{q}},\epsilon_{\mathbf{q}}](\mathbf{r}) = \sum_i^{N_e/2} f_i\Big[\,&\langle\widetilde{\psi}_i| + \lambda_{\mathbf{q}}^*\,\langle\widetilde{\psi}_i{}^{1\lambda_{\mathbf{q}}}| + \lambda_{-\mathbf{q}}^*\,\langle\widetilde{\psi}_i{}^{1\lambda_{-\mathbf{q}}}| + \epsilon_{\mathbf{q}}^*\,\langle\widetilde{\psi}_i{}^{1\epsilon_{\mathbf{q}}}| + \epsilon_{-\mathbf{q}}^*\,\langle\widetilde{\psi}_i{}^{1\epsilon_{-\mathbf{q}}}|\,\Big]\,|\mathbf{r}\rangle \\ &\langle\mathbf{r}|\,\Big[\,|\widetilde{\psi}_i\rangle + \lambda_{\mathbf{q}}\,|\widetilde{\psi}_i{}^{1\lambda_{\mathbf{q}}}\rangle + \lambda_{-\mathbf{q}}\,|\widetilde{\psi}_i{}^{1\lambda_{-\mathbf{q}}}\rangle + \epsilon_{\mathbf{q}}\,|\widetilde{\psi}_i{}^{1\epsilon_{\mathbf{q}}}\rangle + \epsilon_{-\mathbf{q}}\,|\widetilde{\psi}_i{}^{1\epsilon_{-\mathbf{q}}}\rangle\,\Big]\end{aligned} \tag{4.2.7}$$
$$\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}) = \frac{d\widetilde{\rho}}{d\lambda_{-\mathbf{q}}}(\mathbf{r}) = \sum_i^{N_e/2} f_i\Big[\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\mathbf{r}\rangle\,\langle\mathbf{r}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\mathbf{r}\rangle\,\langle\mathbf{r}|\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}\rangle\,\Big] = \sum_i^{N_e/2} 2f_i\,\langle\mathbf{r}|\widetilde{\psi}_i\rangle\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\mathbf{r}\rangle$$

where $f_i = 2$ is the occupation number for the non-spin polarised case and for the last equality we have used Eq. 4.2.5 and the fact that the ground-state wavefunction at Γ $\widetilde{\psi}_i(\mathbf{r})$ is real.

As with the first-order wavefunction (Eq. 4.2.6), we can also define the phase-factorised pseudo-density $\overline{\widetilde{\rho}^{1\lambda_{-q}}}(\mathbf{r})$ in the same manner (but with $\mathbf{q} \to -\mathbf{q}$):

$$\overline{\widetilde{\rho}^{1\lambda_{-q}}}(\mathbf{r}) = e^{i\mathbf{q}\cdot\mathbf{r}}\widetilde{\rho}^{1\lambda_{-q}}(\mathbf{r}) = \sum_i^{N_e/2} 2f_i\,\langle\mathbf{r}|\widetilde{\psi}_i\rangle\,\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|\mathbf{r}\rangle \tag{4.2.8}$$

This will make integration between such local position-dependent first-order quantities significantly easier, especially since it is obvious that:

$$\int f^{1\lambda_{-\mathbf{q}}}(\mathbf{r})g^{1\epsilon_{\mathbf{q}}}(\mathbf{r})\,\mathrm{d}\mathbf{r} = \int \overline{f^{1\lambda_{-\mathbf{q}}}}(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\,\overline{g^{1\epsilon_{\mathbf{q}}}}(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}\mathbf{r} = \int \overline{f^{1\lambda_{-\mathbf{q}}}}(\mathbf{r})\overline{g^{1\epsilon_{\mathbf{q}}}}(\mathbf{r})\,\mathrm{d}\mathbf{r}\ , \quad (4.2.9)$$

which can easily be computed as a sum of Fourier components since the phase-factorised functions share the periodicity of the simulation cell. It is also clear that there is no distinction between factorised and non-factorised quantities at the phonon Gamma point ($\mathbf{q} = \mathbf{0}$).

For the other densities, we need the projector density kernel ρ^{ab} (Eq. 2.2.8), which can also be perturbatively expanded:

$$\begin{aligned}\rho^{ab}[\lambda_{\mathbf{q}},\epsilon_{\mathbf{q}}] = \sum_{i}^{N_e/2} f_i\Big[\langle\widetilde{\psi}_i| + \lambda^*_{-\mathbf{q}}\langle\widetilde{\psi}_i^{\,1\lambda_{-\mathbf{q}}}| + \lambda^*_{\mathbf{q}}\langle\widetilde{\psi}_i^{\,1\lambda_{\mathbf{q}}}| + \epsilon^*_{\mathbf{q}}\langle\widetilde{\psi}_i^{\,1\epsilon_{\mathbf{q}}}| + \epsilon^*_{-\mathbf{q}}\langle\widetilde{\psi}_i^{\,1\epsilon_{-\mathbf{q}}}|\Big]\,|\widetilde{p}^a\rangle \\ \langle\widetilde{p}^b|\widetilde{\psi}_i + \lambda_{-\mathbf{q}}\widetilde{\psi}_i^{\,1\lambda_{-\mathbf{q}}} + \lambda_{\mathbf{q}}\widetilde{\psi}_i^{\,1\lambda_{\mathbf{q}}} + \epsilon_{\mathbf{q}}\widetilde{\psi}_i^{\,1\epsilon_{\mathbf{q}}} + \epsilon_{-\mathbf{q}}\widetilde{\psi}_i^{\,1\epsilon_{-\mathbf{q}}}\rangle \end{aligned} \quad (4.2.10)$$

The first-order equivalent takes the form of:

$$\begin{aligned}\rho^{ab\ 1\lambda_{-\mathbf{q}}} = \sum_n f_n\Big[&\langle\widetilde{\psi}_n^{1\lambda_{\mathbf{q}}}|\widetilde{p}^a\rangle\,\langle\widetilde{p}^b|\widetilde{\psi}_n\rangle + \langle\widetilde{\psi}_n|\widetilde{p}^{a\ 1\lambda_{-\mathbf{q}}}\rangle\,\langle\widetilde{p}^b|\widetilde{\psi}_n\rangle + \\ &\langle\widetilde{\psi}_n|\widetilde{p}^a\rangle\,\langle\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}|\widetilde{\psi}_n\rangle + \langle\widetilde{\psi}_n|\widetilde{p}^a\rangle\,\langle\widetilde{p}^b|\widetilde{\psi}_n^{1\lambda_{-\mathbf{q}}}\rangle\Big] = \\ \sum_n f_n\Big[&\langle\widetilde{p}^a|\psi_n\rangle\,\langle\overline{\widetilde{\psi}_n^{1\lambda_{\mathbf{q}}}}|\widetilde{p}^b_{\mathbf{q}}\rangle + \langle\widetilde{p}^b|\widetilde{\psi}_n\rangle\,\langle\overline{\widetilde{\psi}_n^{1\lambda_{\mathbf{q}}}}|\widetilde{p}^a_{\mathbf{q}}\rangle + \\ &\langle\widetilde{p}^{a\ 1\lambda_{\mathbf{q}}}|\widetilde{\psi}_n\rangle\,\langle\widetilde{\psi}_n|\widetilde{p}^b\rangle + \langle\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}||\widetilde{\psi}_n\rangle\,\langle\widetilde{\psi}_n|\widetilde{p}^a\rangle\Big]\end{aligned} \quad (4.2.11)$$

where $\langle\overline{\widetilde{p}^{a\ 1\lambda_{\mathbf{q}}}}| = |\overline{\widetilde{p}^{a\ 1\lambda_{\mathbf{q}}}}\rangle^\dagger$ are the first-order phase-factorised projectors, whose implementation is described in Sec. 4.3. Moreover,

$$\widetilde{p}^a_{\mathbf{q}}(\mathbf{r}) = e^{-i\mathbf{q}\cdot\mathbf{r}}\,\widetilde{p}^a(\mathbf{r}) = \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\,e^{-i(\mathbf{G}+\mathbf{q})\cdot\mathbf{R}_a}\,\widetilde{p}^a(\mathbf{G}+\mathbf{q}) \quad (4.2.12)$$

are defined to be q-dependent projectors.

At this point, we make a useful approximation: terms defined only in the augmentation sphere are assumed to move with the reference frame of the perturbed atom. Thus, they are not perturbation-dependent when present inside integrals on spheres, but we stress that they are perturbation-dependent in the reference frame of the simulation cell. These

are the partial waves ($\widetilde{\varphi}_a$, φ_a) and the charges ρ_c, $\widetilde{\rho}_c$, ρ_{Z_c}, $\widetilde{\rho}_{Z_c}$. This approximation does not apply to the projectors, as their movement changes their overlap with the PS wavefunction. The same approximation also applies to the multipole moments $\hat{Q}_{ab}(\mathbf{r})$.

Consequently, since $\widetilde{\varphi}_a$ and φ_a are not-perturbation dependent, only the projector density kernel $\rho^{ab}[\lambda_{-\mathbf{q}}, \epsilon\mathbf{q}]$ is perturbation-dependent in the expanded augmentation densities (Eq. 2.2.7):

$$\begin{aligned}\widetilde{\rho}^1[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}](\mathbf{r}) &= \sum_{a,b} \rho^{ab}[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}]\widetilde{\varphi}_a(\mathbf{r})\widetilde{\varphi}_b(\mathbf{r}) \\ \rho^1[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}](\mathbf{r}) &= \sum_{a,b} \rho^{ab}[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}]\varphi_a(\mathbf{r})\varphi_b(\mathbf{r}) \ ,\end{aligned} \tag{4.2.13}$$

with their phase-factorised first-order equivalents being proportional to $\rho^{ab\ 1\lambda_{-\mathbf{q}}}$ (Eq. 4.2.11). The indices a, b run over all the projectors in the simulation cell. The compensation charge on the simulation cell grid has the following form:

$$\hat{\rho}[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}](\mathbf{r}) = \sum_{a,b} \rho^{ab}[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}]\hat{Q}_{ab}[\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}](\mathbf{r}) \tag{4.2.14}$$

Note that on the atom-centered radial grid, according the the previously discussed approximations, the multipole moments $\hat{Q}_{ab}$ would have no perturbation dependence. However, on the simulation cell grid the un-factorised and phase-factorised first-order compensation charge is:

$$\begin{aligned}\hat{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}) &= \sum_{a,b} \rho^{ab\ 1\lambda_{-\mathbf{q}}}\ Q_{ab}(\mathbf{r}) + \sum_{a,b} \rho^{ab}\ \hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}) \\ \overline{\hat{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}) &= \sum_{a,b} \rho^{ab\ 1\lambda_{-\mathbf{q}}}\ Q_{ab,-\mathbf{q}}(\mathbf{r}) + \sum_{a,b} \rho^{ab}\ \overline{\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}) \ ,\end{aligned} \tag{4.2.15}$$

where $\overline{\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r})$ are phase-factorised multipole moments (as described in Sec. 4.3), while

$$Q_{ab,-\mathbf{q}}(\mathbf{r}) = e^{i\mathbf{q}\cdot\mathbf{r}}\ Q_{ab}(\mathbf{r}) \tag{4.2.16}$$

are q-dependent versions of the multipole moments $\hat{Q}_{ab}(\mathbf{r})$. Since the projector density kernel is essentially envisioned as the overlap between the global (throughout all cells) wavefunction with the projectors from only within the original cell, the global q-dependent multipole moments (Eq. 4.2.16) and projectors (Eq. 4.2.12) can simply be considered to be duplicates (with the cell periodicity) of their original cell values. This train of thought

is dependent on the fact that both the projectors and the multipole moments are highly localised within the PAW spheres. We present a formal proof for our statement in Sec. 4.3.3.

With all the necessary densities and their expansions known, we write the perturbation-expanded electronic component of the energy functional (including the Lagrange multipliers enforcing wavefunction-orthonormality):

$$
\begin{aligned}
&E^{\mathrm{EL}}_{\mathrm{TOT}}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] = \sum_{i}^{N_e/2} f_i\big(\,\langle\widetilde{\psi}| + \lambda_{\mathbf{q}}\,\langle\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}| + \epsilon_{-\mathbf{q}}\,\langle\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}|\,\big)\Big[-\frac{1}{2}\nabla^2 + v_{\mathrm{H}}\big[\widetilde{\rho}_{Z_c}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\big]\Big]\big(\,|\widetilde{\psi}_i\rangle + \\
&\lambda_{-\mathbf{q}}\,|\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}\rangle + \epsilon_{\mathbf{q}}\,|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle\,\big) + \sum_{a,b}\rho^{ab}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Bigg\{\int v_{\mathrm{H}}\big[\widetilde{\rho}_{Z_c}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\big]\hat{Q}_{ab}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\;\mathrm{d}\mathbf{r} + \\
&\langle\varphi_a| - \frac{1}{2}\nabla^2 + v_{\mathrm{H}}\big[\rho_{Z_c}\big]\,|\varphi_b\rangle - \langle\widetilde{\varphi}_a| - \frac{1}{2}\nabla^2 + v_{\mathrm{H}}\big[\widetilde{\rho}_{Z_c}\big]\,|\widetilde{\varphi}_b\rangle - \int_{\Omega} v_{\mathrm{H}}\big[\widetilde{\rho}_{Z_c}\big]\hat{Q}_{ab}\;\mathrm{d}\mathbf{r}\Bigg\} + \\
&E_{\mathrm{H}}\Big[\widetilde{\rho}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \hat{\rho}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big] + E_{\mathrm{XC}}\Big[\widetilde{\rho}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \hat{\rho}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \widetilde{\rho}_c[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big] + E_{\mathrm{H}}\Big|_{\Omega}\Big[\rho^1[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big] + \\
&E_{\mathrm{XC}}\Big|_{\Omega}\Big[\rho^1[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \rho_c\Big] - E_{\mathrm{H}}\Big|_{\Omega}\Big[\widetilde{\rho}^1[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \hat{\rho}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big] - E_{\mathrm{XC}}\Big|_{\Omega}\Big[\widetilde{\rho}^1[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \hat{\rho}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}] + \widetilde{\rho}_c\Big] - \\
&\sum_{i,j=1}^{N_e/2}\sum_{\mathbf{k}} f_i\Bigg[\Lambda_{ij\mathbf{k}}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big(\,\langle\widetilde{\psi}_i| + \lambda_{\mathbf{q}}\,\langle\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}| + \lambda_{-\mathbf{q}}\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}| + \langle\epsilon_{-\mathbf{q}}\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}| + \langle\epsilon_{\mathbf{q}}\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}|\,\Big)\hat{S} \\
&\Big(\,|\widetilde{\psi}_{j,\mathbf{k}}\rangle + \lambda_{-\mathbf{q}}\,|\widetilde{\psi}_{j,\mathbf{k}}^{1\lambda_{-\mathbf{q}}}\rangle + \lambda_{\mathbf{q}}\,|\widetilde{\psi}_{j,\mathbf{k}}^{1\lambda_{\mathbf{q}}}\rangle + \epsilon_{\mathbf{q}}\,|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{\mathbf{q}}}\rangle + \epsilon_{-\mathbf{q}}\,|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{-\mathbf{q}}}\rangle\,\Big) - \delta_{ij}\delta_{\mathbf{0},\mathbf{k}}\Bigg] - \mathbf{E}\cdot\mathbf{P}\delta_{\mathbf{q},\mathbf{0}}
\end{aligned}
\tag{4.2.17}
$$

where $\Lambda_{ij\mathbf{k}}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]$ are perturbation-dependent Lagrange multipliers that enforce the orthonormality constraints. Notice that we have generalised the orthogonality constrains to also take into account different k-points $\mathbf{k}$, but one will see in Appendix A that selection rules restrict the value for $\mathbf{k}$ to either $\mathbf{0}$, $\mathbf{q}$ or $-\mathbf{q}$. The term $\mathbf{E}\cdot\mathbf{P}\delta_{\mathbf{q},\mathbf{0}}$ pertains to the interaction of the system with an external electric field $\mathbf{E}$, while $\mathbf{P}$ is the system polarisation vector, explored in section 4.5.

Moreover, we will also need the derivative of general Hartree or exchange-correlation energies with respect to atomic perturbations, for a generic density ρ:

$$
\frac{d}{d\lambda_{-\mathbf{q}}}E_{\mathrm{H/XC}}\Big[\rho[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big] = \int \underbrace{\frac{\delta E_{\mathrm{H/XC}}}{\delta\rho(\mathbf{r})}}_{=v_{\mathrm{H/XC}}[\rho](\mathbf{r})}\;\frac{d\rho[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}](\mathbf{r})}{d\lambda_{-\mathbf{q}}}\;\mathrm{d}\mathbf{r}, \tag{4.2.18}
$$

where we have used the definition of the Hartree/XC potential as the functional derivative of the Hartree/XC energy. For the ground-state case the Hartree potential is linear with

respect to density, but this is not the case for the phase-factorised first-order Hartree potential at non-zero $\mathbf{q}$-wavevector:

$$\overline{v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}\left[\rho[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\right] \neq v_{\mathrm{H}}\left[\overline{\rho^{1\lambda_{-\mathbf{q}}}}\right] \ , \tag{4.2.19}$$

since the Hartree kernel $\frac{1}{|\mathbf{r}-\mathbf{r}'|}$ is non-local. However, if one considers the perturbation without phase-factoring, the linearity is once again recovered:

$$v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}\left[\rho[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\right] = v_{\mathrm{H}}\left[\rho^{1\lambda_{-\mathbf{q}}}\right] \ , \tag{4.2.20}$$

a property which will be very useful in deriving the response-NGWF gradient, as explained in Sec. 4.4.6. Thus, the perturbations of the Hartree and exchange-correlation potential are delicate matters, and are fully described in Sec. 4.3.5 and 4.3.6, respectively. Note that the first-order XC potential is indeed linear with respect to the first-order densities, for local (LDA) and semi-local (GGA) XC functionals.

4.2.2 Second-order expansion and Sternheimer Equation

We derive the second-order energy term by taking the mixed derivative of the expanded PAW energy functional in Eq. 4.2.17. We only show the resulting formula, while the full derivation is available in Appendix A. The variational total second-order energy is:

$$\begin{aligned}
&\frac{d^2E_{\mathrm{TOT}}}{d\lambda_{-\mathbf{q}}d\epsilon_{\mathbf{q}}} = \frac{d^2\left(E_{\mathrm{TOT}}^{EL-1}+E_{\mathrm{TOT}}^{EL-2}+E_{\mathrm{TOT}}^{\text{EL-3}}+E_{\mathrm{TOT}}^{\text{EL-4}}+E_{\mathrm{TOT}}^{\text{EL-5}}\right)}{d\lambda_{-\mathbf{q}}d\epsilon_{\mathbf{q}}} + \frac{\partial^2}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}E_{\mathrm{Ion,core}} = \\
&\sum_{i=1}^{N_e/2} f_i\Bigg\{ \langle\widetilde{\psi}_i\Big|\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}} - \epsilon_i\frac{\partial^2\hat{S}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + 2\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{H}-\epsilon_i\hat{S}\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + 2\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{B}^{\epsilon_{\mathbf{q}}} - \epsilon_i\frac{\partial\hat{S}}{\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle \\
&+ 2\,\langle\widetilde{\psi}_i|\frac{d\hat{H}}{d\lambda_{-\mathbf{q}}} - \epsilon_i\frac{\partial\hat{S}}{\partial\lambda_{-\mathbf{q}}}\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle\Bigg\} + \int v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\widetilde{\rho}_c^{1\epsilon_{\mathbf{q}}}(\mathbf{r})\,\mathrm{d}\mathbf{r} + \\
&\int v_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\frac{\partial^2\widetilde{\rho}_c(\mathbf{r})}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\,\mathrm{d}\mathbf{r} - \sum_{i,j=1}^{N_e/2} f_i\Bigg[\Lambda_{ij\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\Big(\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle + \langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle + \\
&\langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle\Big) + \Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\Big(\langle\widetilde{\psi}_{j,\mathbf{q}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_{j,\mathbf{q}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,-\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\rangle\Big)\Bigg] + \\
&\sum_{i=1}^{N_e/2} 2f_i\frac{\partial\overrightarrow{E}}{\partial\epsilon_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{0}}}|\overrightarrow{\mathcal{D}}|\widetilde{\psi}_i\rangle\,\delta_{\mathbf{q},\mathbf{0}} + \sum_{i=1}^{N_e/2} 2f_i\frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i|\overrightarrow{\mathcal{D}}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{0}}}\rangle\,\delta_{\mathbf{q},\mathbf{0}} + \\
&\frac{\partial^2}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\left[\frac{1}{2}\int_\Omega\int_\Omega\frac{\rho_{Z_c}(\mathbf{r})\rho_{Z_c}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + U\right]
\end{aligned} \tag{4.2.21}$$

The $\widetilde{H}$ term is the pseudo-Hamiltonian from Eq. 2.2.11. The red term contains the second-order core self-interaction and Ewald energy, added to the electronic energy. Note that the self-interaction component in the red term disappears under perturbation, as ρ_{Zc} is rigidly bound to the atom reference frame and strictly localised in the PAW sphere. The green term pertains to the variation of the orthonormality conditions and the associated first-order Lagrange multipliers (to be fixed by a gauge choice). Operators $\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}}$ and $\hat{B}^{\epsilon_{\mathbf{q}}}$ are discussed in Eqs. 4.2.28 and 4.2.29. A keen observer will note that this variational form has a term which is quadratic in first-order wavefunctions. While apparently difficult to calculate, the diagonal components of the variational form will be crucial for the optimisation of the first-order NGWFs, as will be discussed in section 4.4.6.

The two terms before the red one in Eq. 4.2.21 are only present in electric-field perturbations and only at $\mathbf{q} = \mathbf{0}$, meaning that for atomic perturbation they are 0. The first-order polarisation may be expanded by using the considerations at the end of section 4.5. It is crucial to observe that if $\lambda_{-\mathbf{q}}, \epsilon_{\mathbf{q}}$ are electric-field perturbations, then all the perturbations of rigid quantities not related to the electric field itself (such as the core charges, multipole moments, ions, projectors) are 0. The only changes that occur are in the pseudo-wavefunctions.

We can improve the variational second-order energy by employing phase-factorised quantities (using Eq. 4.2.6 for instance), such that all quantities have the periodicity of the underlying lattice. Note that phase factorisation,denoted by the overline bar, is not required for second-order local position-dependent quantities, since they are equivalent to those calculated at $\mathbf{q} = \mathbf{0}$ (Sec. 4.3.4). There are several non-intuitive transformations from the original formulation to the phase-factorised one, but the details are explained in the Sec. 4.3, containing implementation details. Thus, the second-order energy becomes:

$$\frac{d^2E_{TOT}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = \sum_{i=1}^{N/2} f_i \Bigg\{ \langle \widetilde{\psi}_i \Big| \hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}} - \epsilon_i \frac{\partial^2 \hat{S}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}} \Big| \widetilde{\psi}_i \rangle + 2\, \langle \overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}} | \hat{H}_{\mathbf{q},\mathbf{q}} - \epsilon_i \hat{S}_{\mathbf{q},\mathbf{q}} \Big| \overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}} \rangle +$$
$$2\, \langle \widetilde{\psi}_i | (\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}} - \epsilon_i (\hat{S}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}} \Big| \overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}} \rangle + 2\, \langle \overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}} | (\hat{B}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}} - \epsilon_i (\hat{S}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}} \Big| \widetilde{\psi}_i \rangle \Bigg\} +$$
$$\frac{\partial^2 U}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}} + \int \overline{v_{XC}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho} + \hat{\rho} + \widetilde{\rho}_c](\mathbf{r}) \overline{\widetilde{\rho}_c^{1\epsilon_{\mathbf{q}}}}(\mathbf{r})\, \mathrm{d}\mathbf{r} + \int v_{XC}[\widetilde{\rho} + \hat{\rho} + \widetilde{\rho}_c] \frac{\partial^2 \widetilde{\rho}_c}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\, \mathrm{d}\mathbf{r} -$$

$$\sum_{i,j=1}^{N_e/2} f_i \Bigg[\Lambda_{ij\mathbf{q}}^{1\epsilon_\mathbf{q}} \Big(\langle \widetilde{\psi}_i^{1\lambda_\mathbf{q}} | \hat{S}_{\mathbf{q},\mathbf{q}} | \widetilde{\psi}_{j,\mathbf{q}} \rangle + \langle \widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_\mathbf{q}} | \hat{S} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}} | \widetilde{\psi}_{j,\mathbf{q}} \rangle \Big) + \tag{4.2.22}$$

$$\Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}} \Big(\langle \widetilde{\psi}_{j,\mathbf{q}} | \hat{S}_{\mathbf{q},\mathbf{q}} | \widetilde{\psi}_i^{1\epsilon_\mathbf{q}} \rangle + \langle \widetilde{\psi}_{j,\mathbf{q}} | \hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_\mathbf{q}} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \hat{S} | \widetilde{\psi}_{j,-\mathbf{q}}^{1\epsilon_\mathbf{q}} \rangle \Big) \Bigg] +$$

$$\sum_{i=1}^{N_e/2} f_i \left[2 \frac{\partial \mathbf{E}}{\partial \lambda_\mathbf{0}} \cdot \langle \widetilde{\psi}_i | \hat{\mathcal{D}} | \widetilde{\psi}_i^{1\epsilon_\mathbf{0}} \rangle + 2 \frac{\partial \mathbf{E}}{\partial \epsilon_\mathbf{0}} \cdot \langle \widetilde{\psi}_i^{1\lambda_\mathbf{0}} | \hat{\mathcal{D}} | \widetilde{\psi}_i \rangle \right] \delta_{\mathbf{q},\mathbf{0}}$$

In the previous expression we have defined the following ground-state operators (under the assumption that the potentials are all local):

$$\langle \mathbf{r} | \widetilde{H}_{\mathbf{q},\mathbf{q}} | \mathbf{r}' \rangle = e^{-i\mathbf{q}\cdot\mathbf{r}} \langle \mathbf{r} | \widetilde{H} | \mathbf{r}' \rangle e^{i\mathbf{q}\cdot\mathbf{r}'} = e^{-i\mathbf{q}\cdot\mathbf{r}} \langle \mathbf{r} | \Big[-\frac{\nabla^2}{2} \Big] | \mathbf{r}' \rangle e^{i\mathbf{q}\cdot\mathbf{r}'} + \tag{4.2.23}$$

$$\delta_{\mathbf{r},\mathbf{r}'} \Big[v[\widetilde{\rho}_{Z_c}] + v_\mathrm{H}[\widetilde{\rho} + \hat{\rho}] + v_\mathrm{XC}[\widetilde{\rho} + \hat{\rho} + \widetilde{\rho}_c] \Big](\mathbf{r}) + \sum_{a,b} e^{-i\mathbf{q}\cdot\mathbf{r}} \widetilde{p}^a(\mathbf{r}) \; D_{ab} \; \widetilde{p}^b(\mathbf{r}') e^{i\mathbf{q}\cdot\mathbf{r}'}$$

$$\langle \mathbf{r} | \hat{S}_{\mathbf{q},\mathbf{q}} | \mathbf{r}' \rangle = \delta_{\mathbf{r}\mathbf{r}'} + \sum_{a,b} e^{-i\mathbf{q}\cdot\mathbf{r}} \widetilde{p}^a(\mathbf{r}) \; \Big[\langle \varphi_a | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \widetilde{\varphi}_b \rangle \Big] \; \widetilde{p}^b(\mathbf{r}') e^{i\mathbf{q}\cdot\mathbf{r}'} \tag{4.2.24}$$

By using the definition of the q-dependent projector (Eq. 4.2.12) we can rewrite the previous operators as:

$$\langle \mathbf{r} | \widetilde{H}_{\mathbf{q},\mathbf{q}} | \mathbf{r}' \rangle = e^{-i\mathbf{q}\cdot\mathbf{r}} \langle \mathbf{r} | \Big[-\frac{\nabla^2}{2} \Big] | \mathbf{r}' \rangle e^{i\mathbf{q}\cdot\mathbf{r}'} + \delta_{\mathbf{r},\mathbf{r}'} \Big[v[\widetilde{\rho}_{Z_c}] + v_\mathrm{H}[\widetilde{\rho} + \hat{\rho}] + v_\mathrm{XC}[\widetilde{\rho} + \hat{\rho} + \widetilde{\rho}_c] \Big](\mathbf{r}) +$$

$$\sum_{a,b} \langle \mathbf{r} | \widetilde{p}_\mathbf{q}^a \rangle \; D_{ab} \; \langle \widetilde{p}_\mathbf{q}^b | \mathbf{r}' \rangle \tag{4.2.25}$$

$$\langle \mathbf{r} | \hat{S}_{\mathbf{q},\mathbf{q}} | \mathbf{r}' \rangle = \delta_{\mathbf{r}\mathbf{r}'} + \sum_{a,b} \langle \mathbf{r} | \widetilde{p}_\mathbf{q}^a \rangle \; \Big[\langle \varphi_a | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \widetilde{\varphi}_b \rangle \Big] \; \langle \widetilde{p}_\mathbf{q}^b | \mathbf{r}' \rangle \tag{4.2.26}$$

Also, the first-order Lagrange multipliers are defined as:

$$\Lambda_{ij,-\mathbf{q}}^{1\lambda_{-\mathbf{q}}} = \frac{d}{d\lambda_{-\mathbf{q}}} \langle \widetilde{\psi}_{j,-\mathbf{q}} | \widetilde{H} | \widetilde{\psi}_i \rangle = \langle \widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_\mathbf{q}} | \widetilde{H} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \widetilde{H}^{1\lambda_{-\mathbf{q}}} | \widetilde{\psi}_{j,\mathbf{q}} \rangle + \langle \widetilde{\psi}_i^{1\lambda_\mathbf{q}} | \widetilde{H} | \widetilde{\psi}_{j,\mathbf{q}} \rangle =$$

$$\langle \widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_\mathbf{q}} | \widetilde{H} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \widetilde{H}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}} | \widetilde{\psi}_{j,\mathbf{q}} \rangle + \langle \widetilde{\psi}_i^{1\lambda_\mathbf{q}} | \widetilde{H}_{\mathbf{q},\mathbf{q}} | \widetilde{\psi}_{j,\mathbf{q}} \rangle \tag{4.2.27}$$

where we have used the time-reversal symmetry of wavefunctions $\widetilde{\psi}_{j,-\mathbf{q}}(\mathbf{r}) = \widetilde{\psi}_{j,\mathbf{q}}^*(\mathbf{r})$. Lastly, keeping in mind that the the presence or absence of overlines is significant, operators $\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_\mathbf{q}}$ and $(\hat{B}^{\epsilon_\mathbf{q}})_{\mathbf{q},\mathbf{0}}$ are:

$$\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_\mathbf{q}} = \frac{\partial^2 v_\mathrm{H}[\widetilde{\rho}_{Z_c}]}{\partial \lambda_{-\mathbf{q}} \partial \epsilon_\mathbf{q}}(\mathbf{r}) + \sum_{a,b} |\widetilde{p}^a\rangle \langle \widetilde{p}^b| \Bigg(\int \overline{\widetilde{v}_\mathrm{eff}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}') \; \overline{\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}}(\mathbf{r}') \; \mathrm{d}\mathbf{r}' + \int \widetilde{v}_\mathrm{eff}(\mathbf{r}') \; \frac{\partial^2 \hat{Q}_{ab}(\mathbf{r}')}{\partial \lambda_{-\mathbf{q}} \partial \epsilon_\mathbf{q}} \; \mathrm{d}\mathbf{r}' +$$

$$\int \overline{v_\mathrm{H}^{1\epsilon_\mathbf{q}}}[\widetilde{\rho}_{Z_c}](\mathbf{r}') \overline{\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}') \; \mathrm{d}\mathbf{r}' + \int \frac{\partial^2 v_\mathrm{H}[\widetilde{\rho}_{Z_c}]}{\partial \lambda_{-\mathbf{q}} \partial \epsilon_\mathbf{q}}(\mathbf{r}') \hat{Q}_{ab}(\mathbf{r}') \; \mathrm{d}\mathbf{r}' \Bigg) + \tag{4.2.28}$$

$$\sum_{a,b}\left[|\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}\rangle\langle\widetilde{p}^{b}|+|\widetilde{p}^{a}\rangle\langle\widetilde{p}^{b\ 1\epsilon_{-\mathbf{q}}}|\right]\frac{d\left(D^1_{ab}-\widetilde{D}^1_{ab}+\hat{D}_{ab}\right)}{d\lambda_{-\mathbf{q}}}+\sum_{a,b}\left[|\widetilde{p}^{a\ 1\lambda_{-\mathbf{q}}}\rangle\langle\widetilde{p}^{b}|+|\widetilde{p}^{a}\rangle\langle\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}|\right]$$
$$\left(\int\widetilde{v}_{\text{eff}}(\mathbf{r}')\ \hat{Q}^{1\epsilon_{\mathbf{q}}}_{ab}(\mathbf{r}')\ \mathrm{d}\mathbf{r}+\int\overline{v^{1\epsilon_{\mathbf{q}}}_{\text{H}}[\widetilde{\rho}_{Z_c}](\mathbf{r}')}\hat{Q}_{ab,-\mathbf{q}}(\mathbf{r}')\ \mathrm{d}\mathbf{r}'\right)+\sum_{a,b}\left[|\frac{\partial^2\widetilde{p}^{a}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\rangle\langle\widetilde{p}^{b}|+\right.$$
$$\left.|\widetilde{p}^{a\ 1\lambda_{-\mathbf{q}}}\rangle\langle\widetilde{p}^{b\ 1\epsilon_{-\mathbf{q}}}|+|\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}\rangle\langle\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}|+|\widetilde{p}^{a}\rangle\langle\frac{\partial^2\widetilde{p}^{b}}{\partial\lambda_{\mathbf{q}}\partial\epsilon_{-\mathbf{q}}}|\right]\left(D^1_{ab}-\widetilde{D}^1_{ab}+\hat{D}_{ab}\right)$$

and

$$(\hat{B}^{\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}}=\overline{v^{1\epsilon_{\mathbf{q}}}_{\text{H}}[\widetilde{\rho}_{Z_c}](\mathbf{r})}+\sum_{a,b}|\widetilde{p}^{a}_{\mathbf{q}}\rangle\langle\widetilde{p}^{b}|\left(\int\widetilde{v}_{\text{eff}}(\mathbf{r}')\ \hat{Q}^{1\epsilon_{\mathbf{q}}}_{ab}(\mathbf{r}')\ \mathrm{d}\mathbf{r}'+\int\overline{v^{1\epsilon_{\mathbf{q}}}_{\text{H}}[\widetilde{\rho}_{Z_c}](\mathbf{r}')}\ \hat{Q}_{ab,-\mathbf{q}}(\mathbf{r}')\ \mathrm{d}\mathbf{r}'\right)+$$
$$\sum_{a,b}\left[|\overline{\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}}\rangle\langle\widetilde{p}^{b}|+|\widetilde{p}^{a}_{\mathbf{q}}\rangle\langle\widetilde{p}^{b\ 1\epsilon_{-\mathbf{q}}}|\right]\left(D^1_{ab}-\widetilde{D}^1_{ab}+\hat{D}_{ab}\right),\tag{4.2.29}$$

where we used the various PAW definitions from Eqs. 2.2.11 and 2.2.12. Moreover, we define the first-order overlaps

$$(\hat{S}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}=\sum_{a,b}\left[|\widetilde{p}^{a\ 1\lambda_{-\mathbf{q}}}\rangle\langle\widetilde{p}^{b}|+|\widetilde{p}^{a}\rangle\langle\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}|\right]e^{i\mathbf{q}\cdot\mathbf{r}}=\sum_{a,b}\left[|\widetilde{p}^{a\ 1\lambda_{-\mathbf{q}}}\rangle\langle\widetilde{p}^{b}_{\mathbf{q}}|+|\widetilde{p}^{a}\rangle\langle\overline{\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}}|\right]$$
$$(\hat{S}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}}=\sum_{a,b}\left[|\overline{\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}}\rangle\langle\widetilde{p}^{b}|+|\widetilde{p}^{a}_{\mathbf{q}}\rangle\langle\widetilde{p}^{b\ 1\epsilon_{-\mathbf{q}}}|\right],\tag{4.2.30}$$

while the first-order Hamiltonian $(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}$ is thoroughly discussed in Sec. 4.2.4, and $D^{1\lambda_{-\mathbf{q}}}_{ab}=\frac{d}{d\lambda_{-\mathbf{q}}}(\hat{D}_{ab}+D^1_{ab}-\widetilde{D}^1_{ab})$ is also derived in Sec. 4.2.5.

From this variational second-order energy, one can obtain the first-order wavefunction through the associated Euler Lagrange equation (for perturbation $\epsilon_{\mathbf{q}}$ for instance):

$$\frac{\partial}{\partial\,\overline{|\widetilde{\psi}^{1\epsilon_{\mathbf{q}}}_i\rangle}}\left(\frac{d^2}{d\lambda_{-\mathbf{q}}d\epsilon_{\mathbf{q}}}E_{\text{TOT}}\right)=2f_i\,\overline{\langle\widetilde{\psi}^{1\lambda_{\mathbf{q}}}_i|}\left[\widetilde{H}_{\mathbf{q},\mathbf{q}}-\epsilon_i\hat{S}_{\mathbf{q},\mathbf{q}}\right]+2f_i\,\langle\widetilde{\psi}_i|\left[(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}-\epsilon_i(\hat{S}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}\right]+$$
$$2f_i\frac{\partial\mathbf{E}}{\partial\lambda_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i|\,\hat{\mathcal{D}}\ \delta_{\mathbf{q},\mathbf{0}}-\sum_{j=1}^{N_e/2}f_i\Lambda^{1\lambda_{-\mathbf{q}}}_{ij,-\mathbf{q}}\,\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}|}\,\hat{S}_{\mathbf{q},\mathbf{q}}=0\tag{4.2.31}$$

Obviously, the red term is problematic, as it includes the first-order Lagrange multiplier. To eliminate it, we define an out-of-valence-manifold projector $\hat{P}_{c\ \mathbf{q},\mathbf{q}}$ which has the property that $\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}|}\,\hat{S}_{\mathbf{q},\mathbf{q}}\hat{P}_{c\ \mathbf{q},\mathbf{q}}=0$, while also requiring its idempotency $\hat{P}_{c\ \mathbf{q},\mathbf{q}}\hat{P}_{c\ \mathbf{q},\mathbf{q}}=\hat{P}_{c\ \mathbf{q},\mathbf{q}}$. One form that fulfils these purposes is

$$\hat{P}_{c\ \mathbf{q},\mathbf{q}}=\underbrace{\sum_{n\in all}|\overline{\widetilde{\psi}_{n,\mathbf{q}}}\rangle\langle\overline{\widetilde{\psi}_{n,\mathbf{q}}}|}_{\hat{S}^{-1}_{\mathbf{q},\mathbf{q}}}\hat{S}_{\mathbf{q},\mathbf{q}}-\underbrace{\sum_{n\in occ}|\overline{\widetilde{\psi}_{n,\mathbf{q}}}\rangle\langle\overline{\widetilde{\psi}_{n,\mathbf{q}}}|}_{\hat{P}_{v\ \mathbf{q},\mathbf{q}}}\hat{S}_{\mathbf{q},\mathbf{q}}=\hat{1}-\hat{P}_{v\ \mathbf{q},\mathbf{q}},\tag{4.2.32}$$

where $|\overline{\widetilde{\psi}_{n,\mathbf{q}}}\rangle = e^{-i\mathbf{q}\cdot\mathbf{r}}\ |\widetilde{\psi}_{n,\mathbf{q}}\rangle$ is a cell-periodic function derived by phase-factoring the ground-state wavefunction at $\mathbf{q}$.

Thus, by combining Eqs. 4.2.31 and 4.2.32 we obtain:

$$\langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\,[\widetilde{H}_{\mathbf{q},\mathbf{q}} - \epsilon_i\hat{S}_{\mathbf{q},\mathbf{q}}]\hat{P}_{c\ \mathbf{q},\mathbf{q}} = -\langle\widetilde{\psi}_i|\left[(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}} - \epsilon_i(\hat{S}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}} + \left(\frac{\partial\mathbf{E}}{\partial\lambda_\mathbf{0}}\cdot\hat{\mathcal{D}}\right)\delta_{\mathbf{q},\mathbf{0}}\right]\hat{P}_{c\ \mathbf{q},\mathbf{q}}\ , \tag{4.2.33}$$

with i being restricted to the occupied states only.

We can further restrict the solution space by employing a parallel transport gauge through a unitary transformation on the wavefunctions (see Sec. 4.3.1 for details). This gauge enforces the first-order wavefunctions to follow:

$$\langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle = \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\rangle = -\frac{1}{2}\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle\ , \tag{4.2.34}$$

Note that the exact form of the gauge transformation is at the convenience of the user, since the final second-order energies should be invariant with respect to the gauge choice. By applying the conduction projector (Eq. 4.2.32) on $\langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|$, and using Eq. 4.2.34, we obtain:

$$\begin{aligned}\langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\,(\hat{P}_c^\dagger)_{\mathbf{q},\mathbf{q}} = \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| - \sum_j \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle\,\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}| = \\ \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| + \frac{1}{2}\sum_j \langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle\,\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}|\end{aligned} \tag{4.2.35}$$

meaning that

$$\langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| = \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\,(\hat{P}_c^\dagger)_{\mathbf{q},\mathbf{q}} - \frac{1}{2}\langle\widetilde{\psi}_i|\,|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\underbrace{\sum_j |\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle\,\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}|}_{\hat{S}_{\mathbf{q},\mathbf{q}}^{-1}\hat{P}_{v\ \mathbf{q},\mathbf{q}}^\dagger} \tag{4.2.36}$$

Therefore, we can replace $\langle\overline{\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}}|$ from Eq. 4.2.33 with the relation derived in Eq. 4.2.36. The second term of Eq. 4.2.36 can be shown to disappear in this context, further simplifying our deductions:

$$\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}|\,[\widetilde{H}_{\mathbf{q},\mathbf{q}} - \epsilon_i\hat{S}_{\mathbf{q},\mathbf{q}}]\hat{P}_{c\ \mathbf{q},\mathbf{q}} = \underbrace{\langle\widetilde{\psi}_{j,\mathbf{q}}|\,\hat{S}_{\mathbf{q},\mathbf{q}}\hat{P}_{c\ \mathbf{q},\mathbf{q}}}_{0}(\epsilon_{j,\mathbf{q}} - \epsilon_i) = 0 \tag{4.2.37}$$

Thus, the final Sternheimer equation in the parallel transport gauge is:

$$\langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\,(\hat{P}_c^\dagger)_{\mathbf{q},\mathbf{q}}[\widetilde{H}_{\mathbf{q},\mathbf{q}} - \epsilon_i\hat{S}_{\mathbf{q},\mathbf{q}}]\hat{P}_{c\ \mathbf{q},\mathbf{q}} + \tag{4.2.38}$$

$$\langle\widetilde{\psi}_i|\left[(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}-\epsilon_i(\hat{S}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}+\left(\frac{\partial\mathbf{E}}{\partial\lambda_{\mathbf{0}}}\cdot\hat{\mathcal{D}}\right)\delta_{\mathbf{q},\mathbf{0}}\right]\hat{P}_{c\ \mathbf{q},\mathbf{q}}=0$$

In essence, all the terms present in this final equation have the periodicity of the simulation cell, which simplifies its solving. For $\mathbf{q}=\mathbf{0}$, Eq. 4.2.38 matches the equation for the plane-wave PAW case in Abinit[108], except for the additional blue term present in electric-field perturbations; in the case of atomic displacements the blue term is null. For electric field perturbations, the perturbation of the overlap operator in Eq. 4.2.30 is null, while the blue term in the Sternheimer Equation is non-zero. Lastly, the derivations related to the first-order Hamiltonian $(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}$ can be studied in Sec. 4.2.4.

4.2.3 Non-variational second-order energy

One can obtain a non-variational form of the second-order total energy, which presents itself as an inexpensive method of obtaining the final energy terms at the end of the optimisation procedures. Thus, for the converged first-order wavefunction, one can apply $\langle\overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}}|$ to the right of the initial form of the Sternheimer equation (Eq. 4.2.31), obtaining:

$$\begin{aligned}&2f_i\,\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|\widetilde{H}_{\mathbf{q},\mathbf{q}}-\epsilon_i\hat{S}_{\mathbf{q},\mathbf{q}}|\overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}}\rangle+2f_i\,\langle\widetilde{\psi}_i|(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}-\epsilon_i(\hat{S}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}}|\overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}}\rangle+\\ &2f_i\frac{\partial\mathbf{E}}{\partial\lambda_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i|\hat{\mathcal{D}}|\overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}}\rangle\ \delta_{\mathbf{q},\mathbf{0}}=\sum_{j=1}^{N_e/2}f_i\Lambda_{ij,-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\overline{\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}}\rangle=\sum_{j=1}^{N_e/2}f_i\Lambda_{ij,-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\epsilon_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle=\\ &-\frac{1}{2}\sum_{j=1}^{N_e/2}f_i\Lambda_{ij,-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}|\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle\ ,\end{aligned}\tag{4.2.39}$$

where the last equality was enforced by our gauge choice (Eq. 4.2.34). Therefore, by replacing Eq. 4.2.39 into the variational second-order energy (Eq. 4.2.22), we arrive at:

$$\begin{aligned}&\frac{d^2E_{TOT}}{d\lambda_{-\mathbf{q}}\,d\epsilon_{\mathbf{q}}}=\sum_{i=1}^{N/2}f_i\left\{\langle\widetilde{\psi}_i\Big|\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}}-\epsilon_i\frac{\partial^2\hat{S}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle+2\,\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|(\hat{B}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}}-\epsilon_i(\hat{S}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}}\Big|\widetilde{\psi}_i\rangle\right\}+\\ &\frac{\partial^2U}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}+\int\overline{v_{XC}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\overline{\widetilde{\rho}_c^{1\epsilon_{\mathbf{q}}}}(\mathbf{r})\,\mathrm{d}\mathbf{r}+\int v_{XC}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c]\frac{\partial^2\widetilde{\rho}_c}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\,\mathrm{d}\mathbf{r}-\\ &\sum_{i,j=1}^{N_e/2}f_i\left[\Lambda_{ij\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\left(\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle+\langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle+\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle\right)+\Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\langle\overline{\widetilde{\psi}_{j,\mathbf{q}}}|\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle\right]\\ &+\sum_{i=1}^{N_e/2}2f_i\frac{\partial\mathbf{E}}{\partial\epsilon_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{0}}}|\hat{\mathcal{D}}|\widetilde{\psi}_i\rangle\ \delta_{\mathbf{q},\mathbf{0}}\end{aligned}\tag{4.2.40}$$

However, by observing the gauge choice (Eq. 4.2.34), it becomes clear that the green term in the previous equation is actually 0:

$$\begin{aligned}&\overline{\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|}\hat{S}_{\mathbf{q},\mathbf{q}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} + \langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} = \\ &\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} - \frac{1}{2}\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} - \frac{1}{2}\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} = 0\end{aligned} \tag{4.2.41}$$

Consequently, the fully simplified form of the non-variational second order energy becomes:

$$\begin{aligned}\frac{d^2E_{TOT}}{d\lambda_{-\mathbf{q}}\,d\epsilon_{\mathbf{q}}} = &\sum_{i=1}^{N/2} f_i\Bigg\{\langle\widetilde{\psi}_i\Big|\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}} - \epsilon_i\frac{\partial^2\hat{S}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + 2\,\overline{\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|}(\hat{B}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}} - \epsilon_i(\hat{S}^{1\epsilon_{\mathbf{q}}})_{\mathbf{q},\mathbf{0}}\Big|\widetilde{\psi}_i\rangle\Bigg\} + \\ &\int\overline{v_{XC}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\overline{\widetilde{\rho}_c^{1\epsilon_{\mathbf{q}}}}(\mathbf{r})\,d\mathbf{r} + \int v_{XC}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c]\frac{\partial^2\widetilde{\rho}_c}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\,d\mathbf{r} - \\ &\sum_{i,j=1}^{N_e/2} f_i\Lambda_{ij}^{1\lambda_{-\mathbf{q}}}\,\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}\Big|}\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_j\rangle + \sum_{i=1}^{N_e/2} 2f_i\frac{\partial\mathbf{E}}{\partial\epsilon_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i^{1\lambda_0}|\hat{\mathcal{D}}|\widetilde{\psi}_i\rangle\;\delta_{\mathbf{q},\mathbf{0}} + \frac{\partial^2U}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\end{aligned} \tag{4.2.42}$$

while the first-order Lagrange multiplier can be further manipulated to read:

$$\begin{aligned}\Lambda_{ij,-\mathbf{q}}^{1\lambda_{-\mathbf{q}}} = &\underbrace{\langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\widetilde{H}|\widetilde{\psi}_i\rangle}_{\epsilon_i\langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle} + \langle\widetilde{\psi}_i|\widetilde{H}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} + \underbrace{\overline{\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|}\widetilde{H}_{\mathbf{q},\mathbf{q}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle}}_{\epsilon_{j,\mathbf{q}}\overline{\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|}\hat{S}_{\mathbf{q},\mathbf{q}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle}} = \\ &-\frac{1}{2}\epsilon_i\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} + \langle\widetilde{\psi}_i|\widetilde{H}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} - \frac{1}{2}\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle}\,\epsilon_{j,\mathbf{q}}\end{aligned} \tag{4.2.43}$$

where in the underbraces we have used the Schrödinger equation, and the second equality was enabled through the gauge choice. Furthermore, the Schrödinger equation further tells us that $\epsilon_i\langle\widetilde{\psi}_i| = \langle\widetilde{\psi}_i|\,\widetilde{H}\hat{S}^{-1}$ and $\epsilon_{j,\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle = \hat{S}_{\mathbf{q},\mathbf{q}}^{-1}\widetilde{H}_{\mathbf{q},\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle$, which finally leads us to

$$\Lambda_{ij,-\mathbf{q}}^{1\lambda_{-\mathbf{q}}} = -\frac{1}{2}\langle\widetilde{\psi}_i|\widetilde{H}\hat{S}^{-1}\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} + \langle\widetilde{\psi}_i|\widetilde{H}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} - \frac{1}{2}\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\hat{S}_{\mathbf{q},\mathbf{q}}^{-1}\widetilde{H}_{\mathbf{q},\mathbf{q}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle} \tag{4.2.44}$$

Finally, we rewrite the following sum of products containing the first-order Lagrange multipliers:

$$\begin{aligned}\sum_{i,j}^{N_e/2} f_i\Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\,\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}|}\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle = &\sum_{i,j}^{N_e/2} f_i\Bigg\{-\frac{1}{2}\langle\widetilde{\psi}_i|\widetilde{H}\hat{S}^{-1}\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle}\,\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}|}\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \\ &\langle\widetilde{\psi}_i|\widetilde{H}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle}\,\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}|}\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle - \frac{1}{2}\langle\widetilde{\psi}_i|\hat{S}_{\mathbf{0},\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\hat{S}_{\mathbf{q},\mathbf{q}}^{-1}\widetilde{H}_{\mathbf{q},\mathbf{q}}\overline{|\widetilde{\psi}_{j,\mathbf{q}}\rangle}\,\overline{\langle\widetilde{\psi}_{j,\mathbf{q}}|}\hat{S}_{\mathbf{q},\mathbf{0}}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle\Bigg\}\end{aligned} \tag{4.2.45}$$

where we have used the definition for the q-dependent valence projector $\hat{P}_{v\ \mathbf{q},\mathbf{q}}$ in the underbrace of Eq. 4.2.32 and $s = 2$ is the occupation number for the non spin-polarised case.

Crucially, all the self-consistent perturbative terms only depend on the $\lambda_{-\mathbf{q}}$ perturbation (such as $\overline{\widetilde{v}_{\text{eff}}^{\lambda_{-\mathbf{q}}}}$ from $\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}}$ in Eq. 4.2.28), while the $\epsilon_{\mathbf{q}}$ perturbation appears in a non-self-consistent manner (such as in perturbed fixed charges or pseudopotentials). This implies that by determining a single first-order wavefunction ($\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}$), one can afterwards non-variationally build an entire row of the force-constant matrix.

However, it is important to note that this energy form is not suitable for minimisation, for which only the variational form can be used. Under the parallel-gauge constraint the green terms in the variational form (Eq. 4.2.22) disappear, leaving an expression without the first-order Lagrange multipliers.

4.2.4 First-order Hamiltonian

Having appeared frequently in the previous section, the expression for the first-order Hamiltonian operator is:

$$\begin{aligned}(\widetilde{H}^{1\lambda_{-\mathbf{q}}})_{\mathbf{0},\mathbf{q}} =& \overline{v_{\text{H}}^{1\lambda_{-\mathbf{q}}}}\big[\widetilde{\rho}_{Z_c}\big](\mathbf{r}) + \overline{v_{\text{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}) + \overline{v_{XC}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}) + \\ & \sum_{a,b}\Big[\,|\widetilde{p}^{a\ 1\lambda_{-\mathbf{q}}}\rangle\langle\widetilde{p}^{b}_{\mathbf{q}}| + |\widetilde{p}^{a}\rangle\langle\overline{\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}}|\,\Big]\left(\hat{D}_{ab} + D^1_{ab} - \widetilde{D}^1_{ab}\right) + \\ & \sum_{a,b}|\widetilde{p}^{a}\rangle\langle\widetilde{p}^{b}_{\mathbf{q}}|\underbrace{\frac{d}{d\lambda_{-\mathbf{q}}}\left(\hat{D}_{ab} + D^1_{ab} - \widetilde{D}^1_{ab}\right)}_{D_{ab}^{1\lambda_{-\mathbf{q}}}}\,,\end{aligned} \tag{4.2.46}$$

where the first-order phase-factorised potentials are described in Sec. 4.3.5 and 4.3.6. The q-dependent and first-order projectors are also described in Sec. 4.3.3. Therefore the only unknown term of the first-order Hamiltonian is the first-order nonlocal term $D_{ab}^{1\lambda_{-\mathbf{q}}}$, discussed in detail further.

4.2.5 First-order $\mathbf{D_{ab}}$ terms

We first express the general equation for the first-order nonlocal energy terms D_{ab} under a perturbation $\lambda_{-\mathbf{q}}$ (leaving out terms containing perturbations of core charges, partial

waves, or multipole moments on the atom-centered grid):

$$\begin{aligned}
&\frac{d}{d\lambda_{-\mathbf{q}}}\left(\hat{D}_{ab} + D^1_{ab} - \widetilde{D}^1_{ab}\right) = \Big\{ \int v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}\left[\widetilde{\rho}_{Zc}\right](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int v_H\left[\widetilde{\rho}_{Zc}\right](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \\
&\int v_H\left[\hat{\rho}+\widetilde{\rho}\right](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}\left[\widetilde{\rho}+\hat{\rho}\right](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \\
&\int v_{XC}^{1\lambda_{-\mathbf{q}}}\left[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c\right](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int v_{XC}\left[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c\right](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r}\Big\} + \\
&\left\{\int_\Omega v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}\left[\rho^1\right](\mathbf{r})\varphi_a(\mathbf{r})\varphi_b(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int_\Omega v_{XC}^{1\lambda_{-\mathbf{q}}}\left[\rho^1+\rho_c\right](\mathbf{r})\varphi_a(\mathbf{r})\varphi_b(\mathbf{r})\ \mathrm{d}\mathbf{r}\right\} - \\
&\left\{\int_\Omega v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}\left[\widetilde{\rho}^1+\hat{\rho}\right](\mathbf{r})\left(\widetilde{\varphi}_a(\mathbf{r})\widetilde{\varphi}_a(\mathbf{r}) + \hat{Q}_{ab}(\mathbf{r})\right)\ \mathrm{d}\mathbf{r} + \right. \\
&\left.\int_\Omega v_{XC}^{1\lambda_{-\mathbf{q}}}\left[\widetilde{\rho}^1+\hat{\rho}+\widetilde{\rho}_c\right](\mathbf{r})\left(\widetilde{\varphi}_a(\mathbf{r})\widetilde{\varphi}_a(\mathbf{r}) + \hat{Q}_{ab}(\mathbf{r})\right)\ \mathrm{d}\mathbf{r}\right\} = \frac{d}{d\lambda_{-\mathbf{q}}}\hat{D}_{ab} + \frac{d}{d\lambda_{-\mathbf{q}}}D^H_{ab} + \frac{d}{d\lambda_{-\mathbf{q}}}D^{XC}_{ab}
\end{aligned} \tag{4.2.47}$$

In the last equality we have split the first-order nonlocal energies using a similar approach as for the zeroth-order D_{ab}: $\frac{d}{d\lambda_{-\mathbf{q}}}D^H_{ab}$ contains the sphere Hartree terms, and $\frac{d}{d\lambda_{-\mathbf{q}}}D^{XC}_{ab}$ contains the sphere exchange-correlation part. While the previously mentioned terms are computed on radial grids, $\frac{d}{d\lambda_{-\mathbf{q}}}\hat{D}_{ab}$ needs to be computed on the standard grid. Phase factorisation can also be used for the latter term, making it:

$$\begin{aligned}
\frac{d}{d\lambda_{-\mathbf{q}}}\hat{D}_{ab} = &\int \overline{v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}\left[\widetilde{\rho}_{Zc}\right](\mathbf{r})\hat{Q}_{ab,\mathbf{q}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int v_H\left[\widetilde{\rho}_{Zc}\right](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \\
&\int v_H\left[\hat{\rho}+\widetilde{\rho}\right](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int \overline{v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}\left[\widetilde{\rho}+\hat{\rho}\right](\mathbf{r})\hat{Q}_{ab,\mathbf{q}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \\
&\int \overline{v_{XC}^{1\lambda_{-\mathbf{q}}}}\left[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c\right](\mathbf{r})\hat{Q}_{ab,\mathbf{q}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int v_{XC}\left[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c\right](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r}
\end{aligned} \tag{4.2.48}$$

We remind the reader of our implicit approximations: in quantities expressed in the atom's reference frame, terms that are tightly bound to the ion itself have a 0 in first order expansion; this includes the pseudo and all-electron partial waves $(\varphi, \widetilde{\varphi})$, the core and pseudo-core charges $(\rho_c, \widetilde{\rho}_c$ and $\widetilde{\rho}_{Zc}, \widetilde{\rho}_{Zc})$, as well as the multipole moments $(\hat{Q}_{ab})$. However, $\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}$ is non-zero in the $\frac{d}{d\lambda_{-\mathbf{q}}}\hat{D}_{ab}$ terms, since it is computed on the standard grid, in the reference frame of the simulation cell.

Hartree terms

Firstly, one must remember the expression for the zero-order D^H_{ab} terms

$$D^H_{ab} = \langle\varphi_a|v_H[\rho^1]|\varphi_b\rangle - \langle\widetilde{\varphi}_a|v_H[\widetilde{\rho}^1+\hat{\rho}]|\widetilde{\varphi}_b\rangle - \sum_{LM}\int_\Omega v_H[\widetilde{\rho}^1](\mathbf{r})\hat{Q}^{LM}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r} -$$

$$\sum_{LM} \int_{\Omega} v_H[\hat{\rho}](\mathbf{r}) \hat{Q}_{ab}^{LM}(\mathbf{r}) \, \mathrm{d}\mathbf{r} = \sum_{k,l} \rho^{kl} e_{abkl} \ , \tag{4.2.49}$$

where e_{abkl} is a fourth-rank tensor that is already computed in ONETEP, while the multipole moments are defined as:

$$\hat{Q}_{ab}(\mathbf{r}) = \sum_{LM} \hat{Q}_{ab}^{LM} = \sum_{LM} q_{ab}^{LM} \ g_L(r) \ S_{LM}(\hat{\mathbf{r}}) \ , \tag{4.2.50}$$

where q_{ab}^{LM} is once again a pre-calculated quantity, $g_L(r)$ is a radial shape function for angular-momentum L, and $S_{LM}(\hat{\mathbf{r}})$ are real-valued spherical harmonics.

Upon perturbation, only the projector density kernel ρ^{ab} changes, as the quantities related to the fourth-rank tensor ϵ_{abkl} move together with the ion core under perturbation. Thus the first order D_{ab}^{H} (red item in Eq. 4.2.47) is simply expressed as:

$$\frac{d \ D_{ab}^{H}}{d\lambda_{-\mathbf{q}}} = \sum_{k,l} \rho^{kl \ 1\lambda_{-\mathbf{q}}} \ e_{abkl} \ , \tag{4.2.51}$$

with the first-order projector density-kernel being defined in Eq. 4.2.11.

Exchange-correlation (XC) terms

For the first-order XC sphere terms (blue items in Eq. 4.2.47), one needs to first express the XC potential on a radial grid. It is assumed that the density near the atom core is almost spherical, meaning that one can perform a second-order expansion of the density with respect to its $L = 0, M = 0$ channel:

$$v_{XC}[\rho] = v_{XC}\big[\rho_{00}\big] + \big(\rho(\mathbf{r}) - \rho_{00}(\mathbf{r})\big) \frac{dv_{XC}}{d\rho}\big[\rho_{00}\big] + \frac{\big[\rho(\mathbf{r}) - \rho_{00}(\mathbf{r})\big]^2}{2} \frac{d^2 v_{XC}}{d\rho^2}\big[\rho_{00}\big] \tag{4.2.52}$$

where ρ is a generic density on the radial grid. However, using the obvious relation

$$\rho(\mathbf{r}) - \rho_{00}(\mathbf{r}) = \sum_{L>0,M} \rho_{LM}(r) S_{LM}(\hat{\mathbf{r}}) \ , \tag{4.2.53}$$

along with the expression for the L, M channels for the radial v_{XC} potential

$$v_{LM}^{XC}[\rho](r) = \int v_{XC}[\rho](\mathbf{r}) S_{LM}(\hat{\mathbf{r}}) \ \mathrm{d}\hat{\mathbf{r}} \ , \tag{4.2.54}$$

one obtains the final expression for $v^{XC}[\rho]_{LM}(r)$:

$$v_{LM}^{XC}[\rho](r) = \begin{cases} \sqrt{4\pi} v_{XC}\big[\rho_{00}\big] + \frac{1}{2\sqrt{4\pi}} \frac{d^2 v_{XC}}{d\rho^2}\big[\rho_{00}\big] \sum_{L'>0,M'} \left(\rho_{L'M'}(r)\right)^2 \ , \text{ if } L = 0 \\ \rho_{LM}(r) \frac{dv_XC}{d\rho}\big[\rho_{00}\big] + \frac{1}{2} \frac{d^2 v_{XC}}{d\rho^2}[\rho_{00}] \sum_{L'>0,M',L''>0,M''} \rho_{L'M'}(r) \cdot \\ \qquad \cdot \rho_{L''M''}(r) G_{L'M'L''M''}^{LM} \ , \text{ if } L > 0 \end{cases} \tag{4.2.55}$$

The derivatives in Eq. 4.2.55 are calculated using a central finite-difference method (with δ being an infinitesimal parameter):

$$\begin{aligned}\frac{dv_{XC}}{d\rho}[\rho_{00}](r) =& \frac{v_{xc}[\rho_{00}(1+\delta)](r) - v_{xc}[\rho_{00}(1-\delta)](r)}{2\delta\rho_{00}(r)} \\ \frac{d^2v_{XC}}{d\rho^2}[\rho_{00}](r) =& \frac{v_{xc}[\rho_{00}(1+\delta)](r) + v_{xc}[\rho_{00}(1-\delta)](r) - 2v_{xc}[\rho_{00}](r)}{\delta^2\rho_{00}^2(r)}\end{aligned} \tag{4.2.56}$$

However, from Eq. 4.2.47, it is clear that we also need the first-order quantities $\frac{d}{d\lambda_{-\mathbf{q}}}v_{XC}[\rho](\mathbf{r})$ on the radial grid. We denote these them as exchange-correlation kernels $f_{XC}^{\lambda_{-\mathbf{q}}}[\rho](\mathbf{r})$, and we calculate them using a central finite-difference approach:

$$\begin{aligned}f_{XC}^{\lambda_{-\mathbf{q}}}[\rho](\mathbf{r}) =& v_{XC}^{1\lambda_{-\mathbf{q}}}[\rho](\mathbf{r}) = \frac{v_{XC}[\rho+\delta\rho^{1\lambda_{-\mathbf{q}}}](\mathbf{r}) - v_{XC}[\rho-\delta\rho^{1\lambda_{-\mathbf{q}}}](\mathbf{r})}{2\delta} \\ =& \sum_{L,M}\frac{v_{LM}^{XC}[\rho+\delta\rho^{1\lambda_{-\mathbf{q}}}](r) - v_{LM}^{XC}[\rho-\delta\rho^{1\lambda_{-\mathbf{q}}}](r)}{2\delta}S_{LM}(\hat{\mathbf{r}})\end{aligned} \tag{4.2.57}$$

Thus, it is obvious that the L, M-channel component for $f_{XC}^{\epsilon}[\rho](\mathbf{r})$ takes the form:

$$f_{XC\ LM}^{\lambda_{-\mathbf{q}}}[\rho](r) = \frac{v_{LM}^{XC}[\rho+\delta\rho^{1\lambda_{-\mathbf{q}}}](r) - v_{LM}^{XC}[\rho-\delta\rho^{1\lambda_{-\mathbf{q}}}](r)}{2\delta} \tag{4.2.58}$$

Consequently, we can derive a more explicit form for the first-order D_{ab}^{XC} by expanding the blue terms in Eq. 4.2.47:

$$\begin{aligned}&\frac{d}{d\lambda_{-\mathbf{q}}}D_{ab}^{XC} = \sum_{L,M}\int\ f_{XC\ LM}^{\lambda_{-\mathbf{q}}}[\rho^1+\rho_c](r)\ \varphi_{n_al_a}(r)\varphi_{n_bl_b}(r)\ G_{l_am_al_bm_b}^{LM}\ \mathrm{d}r - \\ &\sum_{L,M}\int\int_{\Omega} f_{XC\ LM}^{\lambda_{-\mathbf{q}}}[\widetilde{\rho}^1+\hat{\rho}+\widetilde{\rho}_c](r)S_{LM}(\hat{\mathbf{r}})\Bigg(\frac{\widetilde{\varphi}_{n_al_a}(r)}{r}\frac{\widetilde{\varphi}_{n_bl_b}(r)}{r}S_{l_am_a}(\hat{\mathbf{r}})S_{l_bm_b}(\hat{\mathbf{r}}) + \\ &\sum_{L',M'}q_{ab}^{L'M'}g_{L'}(r)S_{L'M'}(\hat{\mathbf{r}})\Bigg)\ r^2\ \mathrm{d}r\mathrm{d}\hat{\mathbf{r}}\ ,\end{aligned} \tag{4.2.59}$$

where $G_{l_am_al_bm_b}^{LM} = \int S_{LM}(\hat{\mathbf{r}})S_{l_am_a}S_{l_bm_b}(\hat{\mathbf{r}})(\hat{\mathbf{r}})\ \mathrm{d}\hat{\mathbf{r}}$ are Gaunt coefficients, while $\phi_{n_al_a}(r)$ and $\widetilde{\phi}_{n_al_a}(r)$ are the radial components of the all-electron and pseudo partial waves.

By explicitly integrating over all solid angles $\hat{\mathbf{r}}$ and using both the orthogonality of the real-valued spherical harmonics and the definition of the Gaunt coefficients, we obtain the final expression:

$$\frac{d}{d\lambda_{-\mathbf{q}}}D_{ab}^{XC} = \sum_{L,M}\int\ f_{XC\ LM}^{\lambda_{-\mathbf{q}}}[\rho^1+\rho_c](r)\ \varphi_{n_al_a}(r)\varphi_{n_bl_b}(r)\ G_{l_am_al_bm_b}^{LM}\ \mathrm{d}r - \tag{4.2.60}$$

$$\sum_{L,M}\int f_{XC\ LM}^{\lambda-\mathbf{q}}[\widetilde{\rho}^{1}+\hat{\rho}+\widetilde{\rho}_{c}](r)\left[\widetilde{\varphi}_{n_a l_a}(r)\widetilde{\varphi}_{n_b l_b}(r)G^{LM}_{l_a m_a l_b m_b}+r^2\ q^{LM}_{ab}g_L(r)\right]\ \mathrm{d}r$$

However, it is clear from Eqs. 4.2.58 and 4.2.55 that we also need the L,M-channel components for all the first-order densities. In general, these take the form:

$$\left[\rho^{1\lambda-\mathbf{q}}\right]_{LM}(r)=\int_{\Omega}\left[\rho^{1\lambda-\mathbf{q}}\right](r)\ S_{LM}(\hat{\mathbf{r}})\ \mathrm{d}(\hat{\mathbf{r}}) \tag{4.2.61}$$

We know that the first-order sphere densities (expressed on the atom-centered grid) are:

$$\begin{aligned}
\rho^{1\ 1\lambda-\mathbf{q}}(\mathbf{r})&=\sum_{kl}\rho^{kl\ 1\lambda-\mathbf{q}}\frac{\varphi_{n_k l_k}(r)\varphi_{n_l l_l}(r)}{r^2}S_{l_k m_k}S_{l_l m_l}(\hat{\mathbf{r}})\\
\widetilde{\rho}^{1\ 1\lambda-\mathbf{q}}(\mathbf{r})&=\sum_{kl}\rho^{kl\ 1\lambda-\mathbf{q}}\frac{\widetilde{\varphi}_{n_k l_k}(r)\widetilde{\varphi}_{n_l l_l}(r)}{r^2}S_{l_k m_k}S_{l_l m_l}(\hat{\mathbf{r}})\\
\hat{\rho}^{1\lambda-\mathbf{q}}(\mathbf{r})&=\sum_{kl}\rho^{kl\ 1\lambda-\mathbf{q}}\sum_{LM}\hat{Q}^{LM}_{kl}(\mathbf{r})=\sum_{kl,LM}\rho^{kl\ 1\lambda-\mathbf{q}}q^{LM}_{kl}g_L(r)S_{LM}(\hat{\mathbf{r}})\ ,
\end{aligned} \tag{4.2.62}$$

Therefore, by combining Eqs. 4.2.61 and 4.2.62, one obtains the final form for the L,M channel contribution of the sphere densities:

$$\begin{aligned}
\left[\rho^{1\ 1\lambda-\mathbf{q}}\right]_{LM}(r)&=\sum_{i,j}\rho^{ij\ 1\lambda-\mathbf{q}}\frac{\varphi_{n_i l_i}(r)}{r}\frac{\varphi_{n_j l_j}(r)}{r}\ G^{LM}_{l_i m_i l_j m_j}\\
\left[\widetilde{\rho}^{1\ 1\lambda-\mathbf{q}}\right]_{LM}(r)&=\sum_{i,j}\rho^{ij\ 1\lambda-\mathbf{q}}\frac{\widetilde{\varphi}_{n_i l_i}(r)}{r}\frac{\widetilde{\varphi}_{n_j l_j}(r)}{r}\ G^{LM}_{l_i m_i l_j m_j}\\
\left[\hat{\rho}\ \ ^{1\lambda-\mathbf{q}}\right]_{LM}(r)&=\sum_{ij}\rho^{ij\ 1\lambda-\mathbf{q}}q^{LM}_{ij}g_L(r)
\end{aligned} \tag{4.2.63}$$

4.2.6 Ion-ion interaction

Zero-order

In ONETEP, the Ewald (ion-ion) interaction is based on the following equation:

$$\begin{aligned}
U=&\frac{N_{\text{cells}}}{2}\sum_{i,j=1}^{N_{\text{I}}}\sum_{\mathbf{R}}\frac{Z_i Z_j}{|\mathbf{R}_i-\mathbf{R}_j+\mathbf{R}|}\left[1-\text{erf}(\eta|\mathbf{R}_i-\mathbf{R}_j+\mathbf{R}|)\right]+\\
&\frac{2\pi N_{\text{cells}}}{V}\sum_{i,j=1}^{N_{\text{I}}}\sum_{\mathbf{G}\neq\mathbf{0}}\frac{Z_i Z_j}{G^2}e^{-\frac{G^2}{4\eta}}\ e^{i\mathbf{G}\cdot(\mathbf{R}_i-\mathbf{R}_j)}-\frac{\pi}{2V\eta^2}\Big(\sum_{i=1}^{N_{\text{I}}}Z_i\Big)^2-\frac{\eta}{\sqrt{\pi}}\sum_{i=1}^{N_{\text{I}}}Z_i^2\ ,
\end{aligned} \tag{4.2.64}$$

where η is a fixed system-dependent parameter, V is the simulation-cell volume, Z_i is the bare charge of ion i, N_I is the number of atoms in the system, N_cells is the number of cells in the sum, and "erf" denotes the error-function. The red term is denoted as U_REAL, while the blue one is U_RECIP, with the names being reminiscent of the types of sums involved in each term. In U_REAL, the sum over real-space cells and ions must avoid the term $|\mathbf{R}_i - \mathbf{R}_j + \mathbf{R}| = 0$, while in U_RECIP, the $G = 0$ terms must be avoided.

Second-order

The derivation of the second-order Ewald terms is quite lengthy, but it does not involve complicated mathematical artefacts. Therefore, in the following we simply give the final result, split into the real-space and reciprocal-space sums. Note that the weights for each perturbation are encoded in the direction vectors $\hat{u}$.

$$\begin{aligned}\frac{\partial^2 U_\mathrm{REAL}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}} = &\sum_{i,j=1}^{N_I}\sum_{\mathbf{R}}\frac{Z_i Z_j}{2r_{ij\mathbf{R}}^3}(1+e^{i\mathbf{q}\cdot\mathbf{R}})[(\mathbf{u}_{\epsilon(j)}\cdot\mathbf{u}_{\lambda(i)} - \mathbf{u}_{\epsilon(i)}\cdot\mathbf{u}_{\lambda(i)})]\Big[\mathrm{erfc}(\eta r_{ij\mathbf{R}}) + \\ &\frac{2}{\sqrt{\pi}}\eta r_{ij\mathbf{R}} e^{-\eta^2 r_{ij\mathbf{R}}^2}\Big] - \sum_{i,j=1}^{N_I}\sum_{\mathbf{R}}\frac{Z_i Z_j}{2r_{ij\mathbf{R}}^5}(1+e^{i\mathbf{q}\cdot\mathbf{R}})[(\mathbf{u}_{\epsilon(j)}\cdot\mathbf{r}_{ij\mathbf{R}})(\mathbf{u}_{\lambda(i)}\cdot\mathbf{r}_{ij\mathbf{R}}) - \\ &(\mathbf{u}_{\epsilon(i)}\cdot\mathbf{r}_{ij\mathbf{R}})(\mathbf{u}_{\lambda(i)}\cdot\mathbf{r}_{ij\mathbf{R}})]\Big[3\mathrm{erfc}(\eta r_{ij\mathbf{R}}) + \frac{2}{\sqrt{\pi}}\eta r_{ij\mathbf{R}} e^{-\eta^2 r_{ij\mathbf{R}}^2}(3+2\eta^2 r_{ij\mathbf{R}}^2)\Big]\end{aligned} \tag{4.2.65}$$

where $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$, and the sum over $\mathbf{R}$ is such that $r_{ij\mathbf{R}} = |\mathbf{r}_{ij\mathbf{R}}| = |\mathbf{R}_i - \mathbf{R}_j + \mathbf{R}|$ is not 0 at any point. Similarly, the reciprocal-space term is:

$$\begin{aligned}\frac{\partial^2 U_\mathrm{RECIP}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}} = &-\sum_{i,j=1}^{N_I}\sum_{\mathbf{G},\mathbf{G}\neq\mathbf{0}}\frac{4\pi Z_i Z_j}{V\mathbf{G}^2}e^{-\mathbf{G}^2/(4\eta^2)}(\mathbf{u}_{\epsilon(i)}\cdot\mathbf{G})(\mathbf{u}_{\lambda(i)}\cdot\mathbf{G})\,\cos\big[\mathbf{G}\cdot(\mathbf{R}_i - \mathbf{R}_j)\big] + \\ &\sum_{i,j=1}^{N_I}\sum_{\mathbf{G},\mathbf{G}+\mathbf{q}\neq\mathbf{0}}\frac{4\pi Z_i Z_j}{V(\mathbf{G}+\mathbf{q})^2}e^{-(\mathbf{G}+\mathbf{q})^2/(4\eta^2)}\big(\mathbf{u}_{\epsilon(i)}\cdot(\mathbf{G}+\mathbf{q})\big)\big(\mathbf{u}_{\lambda(j)}\cdot(\mathbf{G}+\mathbf{q})\big)e^{-i(\mathbf{G}+\mathbf{q})\cdot(\mathbf{R}_i-\mathbf{R}_j)}\end{aligned} \tag{4.2.66}$$

Finally, $\frac{\partial^2 U}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}$ is simply the sum of the terms in Eqs. 4.2.65 and 4.2.66.

4.3 Implementation details

In this section, we explicitly show how most first-order quantities are calculated either in real or reciprocal space, on the standard grid. In the following, it is important to note

that we use the concept of generalised collective perturbation. For instance, in the case of atom displacements, perturbation ϵ displaces all atoms, with atom I being displaced along the unit vector $\mathbf{u}_{\epsilon(I)}$, itself scaled by a weight $w_{\epsilon(I)}$. In most plane-wave calculations, what we call $\epsilon(I)$ is non-zero only for a single atom I at a time, while the weight $w_{\epsilon(I)}$ is 1.

4.3.1 Gauge choice

We require a gauge choice such that the following is enforced:

$$\langle\widetilde{\psi}_i^{1\lambda_\mathbf{q}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle = \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\rangle \tag{4.3.1}$$

Therefore, this would mean that the first-order orthonormality constraint

$$\frac{d}{d\lambda_{-\mathbf{q}}}\underbrace{\langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle}_{\delta_{ij}\delta_{\mathbf{0},\mathbf{q}}} = \langle\widetilde{\psi}_i^{1\lambda_\mathbf{q}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle + \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\rangle + \langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle = 0 \tag{4.3.2}$$

would imply that

$$\underbrace{\langle\widetilde{\psi}_i^{1\lambda_\mathbf{q}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle}_{\overline{\langle\widetilde{\psi}_i^{1\lambda_\mathbf{q}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle}} = -\frac{1}{2}\underbrace{\langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle}_{\langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle}} \ , \tag{4.3.3}$$

where in the underbraces we have simply used the phase-factorisation definition. To obtain this gauge, we apply a perturbation-dependent unitary transform to the wavefunctions:

$$|\widetilde{\psi}'_{i,\mathbf{k}}[\lambda]\rangle = \sum_{j,\mathbf{p}} U_{i\mathbf{k},j\mathbf{p}}[\lambda]\,|\widetilde{\psi}_{j,\mathbf{p}}[\lambda]\rangle \tag{4.3.4}$$

with the wavefunctions having the form in Eq. 4.2.3, and the unitary transformation being:

$$U_{i\mathbf{k},j\mathbf{p}}[\lambda] = \delta_{ij}\delta_{\mathbf{k},\mathbf{p}} + \lambda_{-\mathbf{q}}U^{1\lambda_{-\mathbf{q}}}_{i\mathbf{k},j\mathbf{p}} + \lambda_\mathbf{q}U^{1\lambda_\mathbf{q}}_{i\mathbf{k},j\mathbf{p}} \tag{4.3.5}$$

Consequently, the zero-order wavefunctions remain unchanged ($\widetilde{\psi}'_{i\mathbf{k}} = \widetilde{\psi}_{i\mathbf{k}}$), while the first-order ones take on the following form:

$$\begin{aligned}
|\widetilde{\psi}'^{\,1\lambda_{-\mathbf{q}}}_{i,\mathbf{k}}\rangle &= |\widetilde{\psi}^{1\lambda_{-\mathbf{q}}}_{i,\mathbf{k}}\rangle + \sum_{j,\mathbf{p}} U^{1\lambda_{-\mathbf{q}}}_{i\mathbf{k},j\mathbf{p}}\,|\widetilde{\psi}_{j,\mathbf{p}}\rangle \\
\langle\widetilde{\psi}'^{\,1\lambda_\mathbf{q}}_{i,\mathbf{k}}| &= \langle\widetilde{\psi}^{1\lambda_\mathbf{q}}_{i,\mathbf{k}}| + \sum_{j,\mathbf{p}} \left(U^{1\lambda_\mathbf{q}}_{i\mathbf{k},j\mathbf{p}}\right)^\dagger \langle\widetilde{\psi}_{j,\mathbf{p}}|
\end{aligned} \tag{4.3.6}$$

Thus, by expanding Eq. 4.3.1 using the definitions in Eqs. 4.3.5 and 4.3.6, we get:

$$\langle\widetilde{\psi}_i^{1\lambda_\mathbf{q}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle + \sum_{n,\mathbf{k}} \underbrace{\langle\widetilde{\psi}_{n\mathbf{k}}|\hat{S}|\widetilde{\psi}_{j\mathbf{q}}\rangle}_{\delta_{nj}\delta_{\mathbf{k},\mathbf{q}}} \left(U_{i\mathbf{0},n\mathbf{k}}^{1\lambda_\mathbf{q}}\right)^\dagger - \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\rangle - \sum_{n,\mathbf{k}} \underbrace{\langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{n,\mathbf{k}}\rangle}_{\delta_{in}\delta_{0\mathbf{k}}} U_{j\mathbf{q},n\mathbf{k}}^{1\lambda_{-\mathbf{q}}} = 0 \ , \tag{4.3.7}$$

resulting into the final condition for the unitary wavefunction transformation

$$U_{j\mathbf{q},i\mathbf{0}}^{1\lambda_{-\mathbf{q}}} - \left(U_{i\mathbf{0},j\mathbf{q}}^{1\lambda_\mathbf{q}}\right)^\dagger = \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\overline{\widetilde{\psi}_{j,\mathbf{q}}}\rangle - \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\rangle \tag{4.3.8}$$

4.3.2 Periodicity of q-dependent projectors

We can deduce the overlap between a NGWF and a q-dependent projector through two different methods, both of them utilising reciprocal space representations. Thus, we first calculate the overlap by using the mathematical definition of the q-dependent projectors, which are still required to have the cell periodicity:

$$\begin{aligned}\langle\phi_\alpha|\widetilde{p}_\mathbf{q}^i\rangle = \langle\phi_\alpha|e^{-i\mathbf{q}\cdot\mathbf{r}}|\widetilde{p}_i\rangle = \int \phi_a^*(\mathbf{r})\ e^{-i\mathbf{q}\cdot\mathbf{r}}\ \widetilde{p}^i(\mathbf{r})\ \mathrm{d}\mathbf{r} = \sum_{\mathbf{G},\mathbf{G}'}\int e^{-i\mathbf{G}\cdot\mathbf{r}}\phi_\alpha(\mathbf{G})^*\ e^{-i\mathbf{q}\cdot\mathbf{r}}\ e^{i\mathbf{G}'\cdot(\mathbf{r}-\mathbf{R}_i)}\\ \widetilde{p}^i(\mathbf{G}')\ \mathrm{d}\mathbf{r} = \sum_{\mathbf{G},\mathbf{G}'}\delta_{\mathbf{G}',\mathbf{q}+\mathbf{G}}\ \phi_\alpha(\mathbf{G})^*\ e^{-i\mathbf{G}'\cdot\mathbf{R}_i}\widetilde{p}^i(\mathbf{G}') = \sum_{\mathbf{G}}\phi_\alpha(\mathbf{G})^* e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{R}_i}\ \widetilde{p}^i(\mathbf{q}+\mathbf{G})\end{aligned} \tag{4.3.9}$$

In the first line we have simply used the Fourier expansion of both $\phi_\alpha(\mathbf{r})$ and $\widetilde{p}^i(\mathbf{r})$, which we know to have the periodicity of the simulation cell. $\mathbf{G},\mathbf{G}'$ are multiples of the reciprocal-space lattice vectors. The phase factor containing $\mathbf{R}_i$ simply translates the projector to the corresponding atom (PAW sphere) location.

For the second method, we must first construct a cell-periodic q-dependent projector by calculating its Fourier expansion coefficients:

$$\begin{aligned}\widetilde{p}_\mathbf{q}^i(\mathbf{G}) = \int e^{-i\mathbf{G}\cdot\mathbf{r}}\widetilde{p}_\mathbf{q}^i(\mathbf{r})\ \mathrm{d}\mathbf{r} = \int e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}\widetilde{p}^i(\mathbf{r})\ \mathrm{d}\mathbf{r} = \sum_{\mathbf{G}'}\int\ e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}\ e^{i\mathbf{G}'\cdot(\mathbf{r}-\mathbf{R}_i)}\ \widetilde{p}^i(\mathbf{G}')\ \mathrm{d}\mathbf{r} =\\ \sum_{\mathbf{G}'}\ \delta_{\mathbf{G}',\mathbf{q}+\mathbf{G}}e^{-i\mathbf{G}'\cdot\mathbf{R}_i}\ \widetilde{p}^i(\mathbf{G}') = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{R}_i}\ \widetilde{p}^i(\mathbf{G}+\mathbf{q})\end{aligned} \tag{4.3.10}$$

Consequently, the second expression for the NGWF-projector overlap is:

$$\langle\phi_\alpha|\widetilde{p}_\mathbf{q}^i\rangle = \int \phi_\alpha^*(\mathbf{r})\ \widetilde{p}_\mathbf{q}^i(\mathbf{r})\ \mathrm{d}\mathbf{r} = \sum_{\mathbf{G},\mathbf{G}'}\int e^{-i\mathbf{G}\cdot\mathbf{r}}\ \phi_\alpha(\mathbf{G})^*\ e^{i\mathbf{G}'\cdot\mathbf{r}}\ \widetilde{p}_\mathbf{q}^i(\mathbf{G}')\ \mathrm{d}\mathbf{r} =$$

$$\sum_{\mathbf{G}} \phi_\alpha(\mathbf{G})^* \; \widetilde{p}^i_{\mathbf{q}}(\mathbf{G}) = \sum_{\mathbf{G}} \phi_\alpha(\mathbf{G})^* e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{R}_i} \; \widetilde{p}^i(\mathbf{q}+\mathbf{G}) \tag{4.3.11}$$

It is clear that the results of Eqs. 4.3.9 and 4.3.11 are identical, meaning that we can certainly think of the q-dependent projectors as having the periodicity of the simulation cell. This exact same approach also applies to the q-dependent multipole moments $\hat{Q}_{ab,\mathbf{q}}(\mathbf{r})$.

4.3.3 First-order (phase-factorised) atom-centred quantities

The first example is that of a projector perturbed by $\epsilon_{\mathbf{q}}$. In the following, I_a denotes the atom of projector a, $\mathbf{R}_{I_a}$ the position of atom I_a from the cell origin, $s(a)$ is the species of projector a, while $\mathbf{R}'$ is a multiple of the real-space lattice vector.

$$\begin{aligned}\overline{\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}}(\mathbf{r}) = e^{-i\mathbf{q}\cdot\mathbf{r}}\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}(\mathbf{r}) = e^{-i\mathbf{q}\cdot\mathbf{r}}\sum_{\mathbf{R}'} e^{i\mathbf{q}\cdot\mathbf{R}'}\frac{\partial}{\partial\epsilon_{\mathbf{R}'}}\widetilde{p}^a(\mathbf{r}-\mathbf{R}') = e^{-i\mathbf{q}\cdot\mathbf{r}}\sum_{\mathbf{R}'} e^{i\mathbf{q}\cdot\mathbf{R}'}\mathbf{u}_{\epsilon(I_a)}\cdot \\ \Big(-\frac{\partial}{\partial\mathbf{r}}\Big)\widetilde{p}^a(\mathbf{r}-\mathbf{R}') = \underbrace{-\sum_{\mathbf{R}'} e^{-i\mathbf{q}\cdot(\mathbf{r}-\mathbf{R}')}\;\mathbf{u}_{\epsilon(I_a)}\cdot\frac{\partial}{\partial\mathbf{r}}\widetilde{p}^a(\mathbf{r}-\mathbf{R}')}\end{aligned} \tag{4.3.12}$$

The term above the brace clearly has the periodicity of the simulation cell, since it is invariant to a translation $\mathbf{r}\to\mathbf{r}+\mathbf{R}$. We can therefore apply a forward and backward Fourier transform upon that term:

$$\overline{\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}}(\mathbf{r}) = -\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\int\sum_{\mathbf{R}'} e^{-i\mathbf{G}\cdot\mathbf{r}'}e^{-i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{R}')}\;\mathbf{u}_{\epsilon(I_a)}\cdot\frac{\partial}{\partial\mathbf{r}'}\widetilde{p}^a(\mathbf{r}'-\mathbf{R}')\;\mathrm{d}\mathbf{r}' \tag{4.3.13}$$

If we perform a change in variable $\mathbf{r}''=\mathbf{r}'-\mathbf{R}'$, we obtain:

$$\begin{aligned}\overline{\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}}(\mathbf{r}) = -\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\sum_{\mathbf{R}'}\underbrace{e^{-i\mathbf{G}\cdot\mathbf{R}'}}_{1}\int e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}''}\;\mathbf{u}_{\epsilon(I_a)}\cdot\frac{\partial}{\partial\mathbf{r}''}\widetilde{p}^a(\mathbf{r}'') = \\ -\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\int e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}''}\;\mathbf{u}_{\epsilon(I_a)}\cdot\frac{\partial}{\partial\mathbf{r}''}\left[\frac{1}{N}\sum_{\mathbf{G}'} e^{i\mathbf{G}'\cdot(\mathbf{r}''-\mathbf{R}_{I_a})}\widetilde{p}^{s(a)}(\mathbf{G}')\right]\mathrm{d}\mathbf{r}''\end{aligned} \tag{4.3.14}$$

In the first line of Eq. 4.3.14 we used the fact that $\mathbf{G}\cdot\mathbf{R}'$ is a multiple of 2π, and in the second line we simply expressed the periodic function $\widetilde{p}^a(\mathbf{r}'')$ as an inverse Fourier transform; as in the previous subsection, the presence of $\mathbf{R}_{I_a}$ in the phase factor displaces the projector atom to the cell origin. Consequently, we reach the final expression:

$$\overline{\widetilde{p}^{a\ 1\epsilon_{\mathbf{q}}}}(\mathbf{r}) = -\sum_{\mathbf{G},\mathbf{G}'} i\mathbf{u}_{\epsilon(I_a)}\cdot\mathbf{G}'\; e^{i\mathbf{G}\cdot\mathbf{r}}\underbrace{\int e^{i(\mathbf{G}'-\mathbf{q}-\mathbf{G})\cdot\mathbf{r}''}\;\mathrm{d}\mathbf{r}''}_{\propto\ \delta_{\mathbf{G}',\mathbf{q}+\mathbf{G}}}\; e^{-i\mathbf{G}'\cdot\mathbf{R}_{I_a}}\widetilde{p}^{s(a)}(\mathbf{G}')\propto \tag{4.3.15}$$

$$- iw_{\epsilon(I_a)}\frac{V}{N}\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\left[(\mathbf{q}+\mathbf{G})_{\epsilon(I_a)}\ e^{-i(\mathbf{q}+\mathbf{G})\mathbf{R}_{I_a}}\ \ \widetilde{p}^{s(a)}(\mathbf{q}+\mathbf{G})\right], \qquad (4.3.15)$$

where $(\mathbf{q}+\mathbf{G})_{\epsilon(I_a)}$ is a component (along the perturbation direction) of vector $\mathbf{q}+\mathbf{G}$. This expression is very easy to calculate by using the current ONETEP subroutines designed for integration by summing Fourier components, with the simple addition that one uses $\mathbf{q}+\mathbf{G}$ instead of just $\mathbf{G}$.

More importantly, this exact approach can be applied to the first-order multipole moments on the standard grid:

$$\overline{\hat{Q}_{ab}^{1\epsilon_{\mathbf{q}}}}(\mathbf{r}) \propto -iw_{\epsilon(I_a)}\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\left[(\mathbf{q}+\mathbf{G})_{\epsilon(I_a)}\ e^{-i(\mathbf{q}+\mathbf{G})\mathbf{R}_{I_a}}\ \ (g_L S_{LM})^{s(a)}(\mathbf{q}+\mathbf{G})\right], \qquad (4.3.16)$$

where $[g_L S_{LM}](\mathbf{q}+\mathbf{G})$ is simply a Fourier coefficient of the product between the atom-centred radially dependent shape function and the associated real spherical harmonic. This can also be applied to the first-order phase-factorised pseudo-core charges $\widetilde{\rho}_c$, with an added sum over atoms:

$$\overline{\widetilde{\rho}_c^{1\epsilon_{\mathbf{q}}}}(\mathbf{r}) \propto \sum_I(-iw_{\epsilon(I)})\frac{V}{N}\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}\left[(\mathbf{q}+\mathbf{G})_{\epsilon(I)}\ e^{-i(\mathbf{q}+\mathbf{G})\mathbf{R}_I}\ \ \widetilde{\rho}_c^{s(I)}(\mathbf{q}+\mathbf{G})\right] \qquad (4.3.17)$$

4.3.4 Second-order projectors

$$\frac{\partial^2}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\widetilde{p}^a(\mathbf{r}) = \sum_{\mathbf{R},\mathbf{R}'} e^{-i\mathbf{q}\cdot\mathbf{R}}\frac{\partial^2}{\partial\lambda_{\mathbf{R}}\partial\epsilon_{\mathbf{R}'}}\widetilde{p}^a(\mathbf{r})\ e^{i\mathbf{q}\cdot\mathbf{R}'} \qquad (4.3.18)$$

However, due to the strict localisation of the PAW sphere, if one applies perturbations $\lambda_{\mathbf{R}}$ and $\epsilon_{\mathbf{R}'}$ in the case of $\mathbf{R}\neq\mathbf{R}'$ (i.e. different cells), the result would be 0. Thus, one is restricted to $\mathbf{R}=\mathbf{R}'$, meaning:

$$\begin{aligned}\frac{\partial^2}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\widetilde{p}^a(\mathbf{r}) &= \sum_{\mathbf{R},\mathbf{R}'} e^{-i\mathbf{q}\cdot\mathbf{R}}\frac{\partial^2}{\partial\lambda_{\mathbf{R}}\partial\epsilon_{\mathbf{R}'}}\widetilde{p}^a(\mathbf{r})\ e^{i\mathbf{q}\cdot\mathbf{R}'}\delta_{\mathbf{R},\mathbf{R}'} = \sum_{\mathbf{R}}\frac{\partial^2}{\partial\lambda_{\mathbf{R}}\partial\epsilon_{\mathbf{R}}}\widetilde{p}^a(\mathbf{r}) \\ &\quad - w_{\epsilon(I_a)}w_{\lambda(I_a)}\frac{V}{N}\sum_{\mathbf{G}}\mathbf{G}_{\epsilon(I_a)}\mathbf{G}_{\lambda(I_a)}e^{i\mathbf{G}\cdot(\mathbf{r}-\mathbf{R}_{I_a})}\ \widetilde{p}^{s(a)}(\mathbf{G}) \qquad (4.3.19)\end{aligned}$$

where to obtain the second line we have used an almost identical approach as in Eq. 4.3.15, with the difference that $\mathbf{q}=0$. This implies that second-order rigid perturbations need to be calculated only once, at $\mathbf{q}=0$. As before, this method can be duplicated in the case of core (and pseudo-core) charges, multipole moments and other quantities that are rigidly perturbed on the standard grid.

4.3.5 First-order (phase-factorised) Hartree potential

The most accessible method of obtaining the phase-factorised first-order Hartree potential is to manipulate its Fourier coefficients. Therefore:

$$\begin{aligned}\overline{v_H^{1\epsilon_{\mathbf{q}}}[\rho]}(\mathbf{r}) =& e^{-i\mathbf{q}\cdot\mathbf{r}} v_H^{1\epsilon_{\mathbf{q}}}[\rho](\mathbf{r}) = e^{-i\mathbf{q}\cdot\mathbf{r}} \frac{d}{d\epsilon_{\mathbf{q}}} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\,\mathrm{d}\mathbf{r}' = e^{-i\mathbf{q}\cdot\mathbf{r}} \int \frac{\rho^{1\epsilon_{\mathbf{q}}}(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\,\mathrm{d}\mathbf{r}' = \\ & e^{-i\mathbf{q}\cdot\mathbf{r}} \int \frac{e^{i\mathbf{q}\cdot\mathbf{r}'}\,\overline{\rho^{1\epsilon_{\mathbf{q}}}}(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\,\mathrm{d}\mathbf{r}' = \int \frac{e^{i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}}{|\mathbf{r}'-\mathbf{r}|} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}'}\,\overline{\rho^{1\epsilon_{\mathbf{q}}}}(\mathbf{G})\,\mathrm{d}\mathbf{r} = \\ & \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \underbrace{\int \frac{e^{i\mathbf{q}\cdot\Delta\mathbf{r}}}{|\Delta\mathbf{r}|}\, e^{i\mathbf{G}\cdot\Delta\mathbf{r}}\,\mathrm{d}\Delta\mathbf{r}}_{\frac{4\pi}{(\mathbf{G}+\mathbf{q})^2}}\,\overline{\rho^{1\epsilon_{\mathbf{q}}}}(\mathbf{G}) = \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \underbrace{\frac{4\pi\overline{\rho^{1\epsilon_{\mathbf{q}}}}(\mathbf{G})}{(\mathbf{G}+\mathbf{q})^2}}_{\overline{v_H^{1\epsilon_{\mathbf{q}}}}[\rho](\mathbf{G})}\end{aligned} \tag{4.3.20}$$

This expression is very similar to what one would use to calculate the ground-state Hartree potential, with the difference that $\mathbf{q} \neq 0$ and instead of $\overline{\rho^{1\epsilon_{\mathbf{q}}}}(\mathbf{G})$ one would have $\rho(\mathbf{G})$. It is worth noting the term in which $\mathbf{G}+\mathbf{q}=\mathbf{0}$ produces a divergence and is therefore skipped.

4.3.6 First-order (phase-factorised) XC potential

For $\mathbf{q}=\mathbf{0}$ (where phase factorisation is not relevant), we express the first-order exchange-correlation potential as:

$$v_{\mathrm{XC}}^{1\lambda_{\mathbf{0}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}) = \int \underbrace{\frac{\delta v_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})}{\delta\big(\widetilde{\rho}(\mathbf{r}')+\hat{\rho}(\mathbf{r}')+\widetilde{\rho}_c(\mathbf{r}')\big)}}_{f_{\mathrm{XC}}(\mathbf{r},\mathbf{r}')} \Big(\widetilde{\rho}^{1\lambda_{\mathbf{0}}}(\mathbf{r}')+\hat{\rho}^{1\lambda_{\mathbf{0}}}(\mathbf{r}')+\widetilde{\rho}_c^{1\lambda_{\mathbf{0}}}(\mathbf{r}')\Big)\,\mathrm{d}\mathbf{r}' \tag{4.3.21}$$

We note that this form does not apply for $\mathbf{q}\neq\mathbf{0}$. The analytical form of the exchange-correlation kernel $f_{XC}(\mathbf{r},\mathbf{r}')$ would involve modifying all the currently-implemented XC functionals to contain the derivative of the potential with respect to the density. Moreover, such an analytical formula would produce an irregular behaviour for small densities. Therefore, under the assumption that the XC potential is local ($f_{\mathrm{XC}}(\mathbf{r},\mathbf{r}') = f_{\mathrm{XC}}(\mathbf{r},\mathbf{r}')\delta_{\mathbf{r},\mathbf{r}'}$) we adapt the approach of T. Zuehlsdorff [109] to use a central finite-differencing approach for the first-order XC potential:

$$v_{\mathrm{XC}}^{1\lambda_{\mathbf{0}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}) = \frac{1}{2\delta}\Big\{ v_{\mathrm{XC}}\big[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c+\delta\big(\widetilde{\rho}^{1\lambda_{\mathbf{0}}}+\hat{\rho}^{1\lambda_{\mathbf{0}}}+\widetilde{\rho}_c^{1\lambda_{\mathbf{0}}}\big)\big](\mathbf{r}) - \tag{4.3.22}$$

$$v_{\mathrm{XC}}\big[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c-\delta\big(\widetilde{\rho}^{1\lambda_{\mathbf{0}}}+\hat{\rho}^{1\lambda_{\mathbf{0}}}+\widetilde{\rho}_c^{1\lambda_{\mathbf{0}}}\big)\big](\mathbf{r})$$

This locality approximation is suitable for LDA and GGAs without Hartree-Fock exchange.

As a last note, we have used the term $\overline{v_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}$ in Eqs. 4.2.28 and 4.2.29. In this context, this is now defined to be:

$$\overline{v_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r})=\overline{v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}_{Z_c}]+\overline{v_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r})+\overline{v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}) \tag{4.3.23}$$

4.3.7 Integrals

In many situations we have integrals of the form $\int A^*(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\ B(\mathbf{r})\ \mathrm{d}\mathbf{r}$, where $A(\mathbf{r})$ and $B(\mathbf{r})$ have the cell periodicity, and one can also have $\mathbf{q}=\mathbf{0}$ or $\mathbf{q}\rightarrow-\mathbf{q}$. These integrals are best calculated in reciprocal space, as follows:

$$\begin{aligned}\int A^*(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\ B(\mathbf{r})\ \mathrm{d}\mathbf{r}=\sum_{\mathbf{G},\mathbf{G}'}\int\ e^{-i\mathbf{G}\cdot\mathbf{r}}A^*(\mathbf{G})\ e^{-i\mathbf{q}\cdot\mathbf{r}}e^{i\mathbf{G}'\cdot\mathbf{r}}\ B(\mathbf{G}')\ \mathrm{d}\mathbf{r}=\\ \sum_{\mathbf{G},\mathbf{G}'}\delta_{\mathbf{G}',\mathbf{q}+\mathbf{G}}\ A^*(\mathbf{G})\ B(\mathbf{G}')=\sum_{\mathbf{G}}A^*(\mathbf{G})B(\mathbf{q}+\mathbf{G})\ ,\end{aligned} \tag{4.3.24}$$

where if $A(\mathbf{r})$ is real, one has $A^*(\mathbf{G})=A(-\mathbf{G})$

4.4 Adapting to NGWF format

4.4.1 Preliminaries

Since ONETEP uses a localised orbital (NGWF) basis, we adopt the same framework to the linear-response phonons formalism. In doing so, we shall depart significantly from any other linear-response implementation, treading into novel territory. Firstly, we cannot use the eigenvalues themselves (in Eq. 4.2.38), as their calculation would ruin the linear-scaling operation. We thus use the Schrödinger equation to express:

$$\epsilon_i\,|\widetilde{\psi}_i\rangle=\hat{S}^{-1}\widetilde{H}\,|\widetilde{\psi}_i\rangle \tag{4.4.1}$$

Furthermore, it will be useful to redefine the out-of-valence projector (Eq. 4.2.32) as a product between a "conduction" density kernel and the overlap operator:

$$\hat{P}_{c\ \mathbf{q},\mathbf{q}} = \hat{K}_{c\ \mathbf{q},\mathbf{q}} \hat{S}_{\mathbf{q},\mathbf{q}} \ , \tag{4.4.2}$$

with the valence equivalent being:

$$\hat{P}_{v\ \mathbf{q},\mathbf{q}} = \sum_j |\widetilde{\psi}_{j,\mathbf{q}}\rangle \langle\widetilde{\psi}_{j,\mathbf{q}}| \hat{S}_{\mathbf{q},\mathbf{q}} = \hat{K}_{v\ \mathbf{q},\mathbf{q}} \hat{S}_{\mathbf{q},\mathbf{q}} \ , \tag{4.4.3}$$

We note that in ONETEP, the valence NGWFs ($|\phi_\alpha\rangle$) are optimised to represent only the valence manifold, but they cannot guarantee an accurate representation of higher-energy states. Therefore, a separate set of "first-order" NGWFs is needed to describe the optimal subspace of perturbed states. Henceforth, the first-order and valence NGWFs are $\{\delta\phi\}$ and $\{\phi\}$, respectively. While it seems intuitive to represent K_c only in the first-order NGWF basis $|\delta\phi_\alpha\rangle$, this scheme would not ensure the idempotency of K_c (similarly to the valence density kernel in Eq. 2.1.6). The solution [109] for this is to use a joint basis $\{\theta\}$ consisting of the union between the response and valence NGWFs $\{\delta\phi\} \cup \{\phi\}$.

The first step is to define the inverse q-dependent overlap operator (θ superscript implies a joint-NGWF representation) as:

$$\hat{S}^{-1}_{\mathbf{q},\mathbf{q}} = |\theta_\alpha\rangle \left(S^{\theta-1}_{\mathbf{q},\mathbf{q}}\right)^{\alpha\beta} \langle\theta_\beta| \ , \tag{4.4.4}$$

with $S^{\theta-1}_{\mathbf{q},\mathbf{q}}$ being the inverse matrix of $S^{\theta}_{\mathbf{q},\mathbf{q}}$. We can test that this is indeed valid by applying $\hat{S}^{-1}_{\mathbf{q},\mathbf{q}}\hat{S}_{\mathbf{q},\mathbf{q}}$ to a NGWF $|\theta_\gamma\rangle$:

$$\hat{S}^{-1}_{\mathbf{q},\mathbf{q}}\hat{S}_{\mathbf{q},\mathbf{q}} |\theta_\gamma\rangle = |\theta_\alpha\rangle \left(S^{\theta-1}_{\mathbf{q},\mathbf{q}}\right)^{\alpha\beta} \langle\theta_\beta| \hat{S}_{\mathbf{q},\mathbf{q}} |\theta_\gamma\rangle = |\theta_\alpha\rangle \underbrace{\left(S^{\theta-1}_{\mathbf{q},\mathbf{q}}\right)^{\alpha\beta} \left(S^{\theta}_{\mathbf{q},\mathbf{q}}\right)_{\beta\gamma}}_{\delta^\alpha{}_\gamma} = |\theta_\gamma\rangle \ , \tag{4.4.5}$$

making it clear that $\hat{S}^{-1}_{\mathbf{q},\mathbf{q}}\hat{S}_{\mathbf{q},\mathbf{q}} = \hat{1}$. From Eqs. 4.4.2 and 4.2.32, we can deduce that the operator $\hat{K}_{c\ \mathbf{q},\mathbf{q}}$ has this form:

$$\hat{K}_{c\ \mathbf{q},\mathbf{q}} = \hat{S}^{-1}_{\mathbf{q},\mathbf{q}} \Big[\hat{S}_{\mathbf{q},\mathbf{q}} - \hat{S}_{\mathbf{q},\mathbf{q}} \sum_k |\widetilde{\psi}_{k,\mathbf{q}}\rangle \langle\widetilde{\psi}_{k,\mathbf{q}}| \hat{S}_{\mathbf{q},\mathbf{q}}\Big] \hat{S}^{-1}_{\mathbf{q},\mathbf{q}} \tag{4.4.6}$$

Therefore, by expanding the inverse overlap (as in Eq. 4.4.4) and density (Eq. 2.1.4) operators, one can obtain the q-dependent out-of-valence projector in NGWF representation:

$$\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}} = \left(\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\right)^{-1} - \left(\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\right)^{-1} \mathbf{S}^{\theta\phi}_{\mathbf{q},\mathbf{q}} \mathbf{K}_{\mathbf{q},\mathbf{q}} \mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}} \left(\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\right)^{-1} \ , \tag{4.4.7}$$

where we calculate the valence density kernel at $\mathbf{q}$ ($\mathbf{K_{q,q}}$) in a non-selfconsistent manner from quantities (such as NGWFs or q-dependent Hamiltonian) already obtained during the ground state calculation at Γ. The essential point about this kernel is that it should obey the idempotency condition:

$$\mathbf{K_{q,q}S_{q,q}K_{q,q}} = \mathbf{K_{q,q}} \tag{4.4.8}$$

Moreover, the conduction-projector operator itself is expressed as:

$$\hat{K}_{c\ \mathbf{q,q}} = |\theta_\alpha\rangle \left(K^{\theta}_{c\ \mathbf{q,q}}\right)^{\alpha\beta} \langle\theta_\beta| \ , \tag{4.4.9}$$

where from now on we assume an implicit summation over Greek indices, and the lack of any $\mathbf{q}$ index implies that the quantity is calculated at $\mathbf{q} = \mathbf{0}$. An identical approach can be taken for the valence projector, yielding:

$$\mathbf{K}^{\theta}_{v\ \mathbf{q,q}} = \left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}}\mathbf{S}^{\phi\theta}_{\mathbf{q,q}}\left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1} \ , \tag{4.4.10}$$

with:

$$\hat{K}_{v\ \mathbf{q,q}} = |\theta_\alpha\rangle \left(K^{\theta}_{v\ \mathbf{q,q}}\right)^{\alpha\beta} \langle\theta_\beta| \ , \tag{4.4.11}$$

First, let us check that the "out-of-valence" kernel projects out of the valence manifold at the $\mathbf{q}$ wavevector:

$$\begin{aligned}\mathbf{K}^{\theta}_{c\ \mathbf{q,q}}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}} =& \Big[\left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1} - \left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}}\mathbf{S}^{\phi\theta}_{\mathbf{q,q}}\left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\Big]\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}} = \\ &\left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}} - \left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\underbrace{\mathbf{K_{q,q}}\mathbf{S}^{\phi\theta}_{\mathbf{q,q}}\left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}}}_{\mathbf{K_{q,q}}\mathbf{S_{q,q}}\mathbf{K_{q,q}}} = \\ &\left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}} - \left(\mathbf{S}^{\theta}_{\mathbf{q,q}}\right)^{-1}\mathbf{S}^{\theta\phi}_{\mathbf{q,q}}\mathbf{K_{q,q}} = \mathbf{O}\end{aligned} \tag{4.4.12}$$

where for the last equality we have used the idempotency of $\mathbf{K_{q,q}}$ (Eq. 4.4.8) on the braced terms. Similarly, it can be proven that the "out-of-valence" kernel is also idempotent, i.e. $\mathbf{K}^{\theta}_{c\ \mathbf{q,q}}\mathbf{S}^{\theta}_{\mathbf{q,q}}\mathbf{K}^{\theta}_{c\ \mathbf{q,q}} = \mathbf{K}^{\theta}_{c\ \mathbf{q,q}}$, meaning that our chosen NGWF-representation for this projector (Eq. 4.4.7) is indeed valid.

For simplicity, quantities in a purely valence (ϕ) representation have no θ or ϕ index, but the superscript $\theta\phi$ suggests a combined joint-valence representation:

$$\left(S^{\theta\phi}_{\mathbf{q,q}}\right)_{\alpha\beta} = \langle\theta_\alpha|\,\hat{S}_{\mathbf{q,q}}\,|\phi_\beta\rangle = \int \theta^*_\alpha(\mathbf{r})\phi_\beta(\mathbf{r})\ \mathrm{d}\mathbf{r} + \sum_{a,b}\langle\theta_\alpha|\widetilde{p}^a_{\mathbf{q}}\rangle\, O_{ab}\,\langle\widetilde{p}^b_{\mathbf{q}}|\phi_\beta\rangle \ , \tag{4.4.13}$$

The next quantity of interest from Eq. 4.2.38 is $|\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}\rangle \langle\widetilde{\psi}_i| = \left(|\widetilde{\psi}_i\rangle \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| \right)^\dagger$. Since for the ground-state one has $\sum_i |\widetilde{\psi}_i\rangle \langle\widetilde{\psi}_i| = |\phi_\alpha\rangle K^{\alpha\beta} \langle\phi_\beta|$, we can define a response kernel $P^{1\lambda_\mathbf{q}\ \theta\phi}$ which describes the changes of the molecular orbital coefficients under perturbation:

$$\sum_i^{N/2} |\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}\rangle \langle\widetilde{\psi}_i| = |\theta_\alpha\rangle \left(P^{1\lambda_\mathbf{q}\ \theta\phi}\right)^{\alpha\beta} \langle\phi_\beta| \quad \text{and} \quad \sum_i^{N/2} |\widetilde{\psi}_i\rangle \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| = |\phi_\beta\rangle \left(P^{1\lambda_\mathbf{q}\ \theta\phi}\right)^{\dagger\ \alpha\beta} \langle\theta_\alpha| \tag{4.4.14}$$

Since $\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}$ has been explicitly defined to have the periodicity of the simulation cell, it is reasonable to represent it using joint NGWFs (θ) which also have the periodicity of the simulation cell. The spin factor (2 for spin-unpolarised systems) will always be explicitly provided separately. It is intuitive to express the response kernel in a mixed joint-valence representation, as it represents the "transition" of a ground-state valence electron into an excited state induced by a perturbation. A similar approach is used in the ONETEP linear-response TDDFT module [109].

As a last note, the only contra-variant quantities that we will need are the density kernel terms $K^{\alpha\beta}$, $\left(K_{c\ \mathbf{q},\mathbf{q}}^{\theta}\right)^{\alpha\beta}$, $\left(K_{v\ \mathbf{q},\mathbf{q}}^{\theta}\right)^{\alpha\beta}$, and $\left(P^{1\lambda_\mathbf{q}\ \theta\phi}\right)^{\alpha\beta}$, and the inverse overlap matrices $\left(\hat{S}^{-1}\right)^{\alpha\beta}$, $\left(\hat{S}_{\mathbf{q},\mathbf{q}}^{-1}\right)^{\alpha\beta}$, and $\left(\hat{S}_{\mathbf{q},\mathbf{q}}^{\theta\ -1}\right)^{\alpha\beta}$.

4.4.2 Gauge choice constraint

The previously derived Sternheimer equation (Eq. 4.2.38) implicitly assumes that the response wavefunction obeys the parallel transport gauge (Eq. 4.2.36). Therefore, we must also explicitly represent the gauge choice in NGWF format. To make use of the definitions in the previous subsection, some steps need to be applied to Eq. 4.2.36:

1. Apply $\hat{S}_{\mathbf{q},\mathbf{q}} |\theta_\beta\rangle$ to the right
2. Apply $\langle\phi_\alpha|\widetilde{S}|\widetilde{\psi}_i\rangle$ to the left
3. Sum over i, the wavefunction band index

The resulting equation, in which we have color-coded terms related to the response wavefunction (red) and the unperturbed wavefunction (blue), is thus:

$$\sum_i \langle\phi_\alpha|\hat{S}|\widetilde{\psi}_i\rangle \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| \hat{S}_{\mathbf{q},\mathbf{q}} |\theta_\beta\rangle - \sum_i \langle\phi_\alpha|\hat{S}|\widetilde{\psi}_i\rangle \langle\overline{\widetilde{\psi}_i^{1\lambda_\mathbf{q}}}| \underbrace{\hat{P}_{c\ \mathbf{q},\mathbf{q}}^{\dagger} \hat{S}_{\mathbf{q},\mathbf{q}}}_{\hat{S}_{\mathbf{q},\mathbf{q}} \hat{K}_{c\ \mathbf{q},\mathbf{q}} \hat{S}_{\mathbf{q},\mathbf{q}}} |\theta_\beta\rangle +$$

$$\frac{1}{2}\sum_i \langle\phi_\alpha|\hat{S}|\widetilde{\psi}_i\rangle\,\langle\widetilde{\psi}_i|\,\hat{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\hat{K}_{v\ \mathbf{q},\mathbf{q}}\hat{S}_{\mathbf{q},\mathbf{q}}\,|\theta_\beta\rangle = \tag{4.4.15}$$

$$\sum_i \langle\phi_\alpha|\hat{S}|\widetilde{\psi}_i\rangle\,\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\theta_\beta\rangle - \sum_i \langle\phi_\alpha|\hat{S}|\widetilde{\psi}_i\rangle\,\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|\hat{S}_{\mathbf{q},\mathbf{q}}|\theta_\eta\rangle\,(K^\theta_{c\ \mathbf{q},\mathbf{q}})^{\eta\rho}\left(S^\theta_{\mathbf{q},\mathbf{q}}\right)_{\rho\beta} +$$

$$\frac{1}{2}\sum_i \langle\phi_\alpha|\hat{S}|\widetilde{\psi}_i\rangle\,\langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}|\theta_\eta\rangle\,(K^\theta_{v\ \mathbf{q},\mathbf{q}})^{\eta\rho}\left(S^\theta_{\mathbf{q},\mathbf{q}}\right)_{\rho\beta} =$$

$$S_{\alpha\gamma}\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \gamma\delta}\left(S^\theta_{\mathbf{q},\mathbf{q}}\right)_{\delta\beta} - S_{\alpha\gamma}\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \gamma\delta}\left(S^\theta_{\mathbf{q},\mathbf{q}}\right)_{\delta\eta}\left(K^\theta_{c\ \mathbf{q},\mathbf{q}}\right)^{\eta\rho}\left(S^\theta_{\mathbf{q},\mathbf{q}}\right)_{\rho\beta} +$$

$$\frac{1}{2}S_{\alpha\gamma}K^{\gamma\delta}\left(S^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\right)_{\delta\eta}\left(K^\theta_{v\ \mathbf{q},\mathbf{q}}\right)^{\eta\rho}\left(S^\theta_{\mathbf{q},\mathbf{q}}\right)_{\rho\beta} = 0$$

where $\left(S^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\right)_{\delta\eta} = \langle\phi_\delta|\hat{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}|\theta_\eta\rangle$. By applying $\mathbf{S}^{-1}$ to the left and $\mathbf{S}^{\theta\ -1}_{\mathbf{q},\mathbf{q}}$ to the right of Eq. 4.4.15, while requiring that the response kernel is invariant to a right-hand side projection on the valence manifold (ensuring the orthonormality of the valence wavefunctions in Eq. 4.4.14), we obtain the matrix-representation of the gauge choice in the NGWF formalism:

$$\left(\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^\dagger = \mathbf{KS}\left(\underline{\mathbf{P}}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^\dagger\mathbf{S}^\theta_{\mathbf{q},\mathbf{q}}\mathbf{K}^\theta_{c\ \mathbf{q},\mathbf{q}} - \frac{1}{2}\underbrace{\mathbf{KSK}}_{\mathbf{K}}\mathbf{S}^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\mathbf{K}^\theta_{v\ \mathbf{q},\mathbf{q}}\,, \tag{4.4.16}$$

where $\mathbf{KSK} = \mathbf{K}$ is the idempotency condition for the valence density kernel at Γ (Eq. 2.1.6). Note the difference between $\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}$ and $\underline{\mathbf{P}}^{1\lambda_{\mathbf{q}}\ \theta\phi}$: as in the case of the valence density-kernel, we first optimise an auxiliary response kernel $\underline{\mathbf{P}}^{1\lambda_{\mathbf{q}}\ \theta\phi}$ that is afterwards forced to obey the gauge constraint through Eq. 4.4.16, thus finally obtaining the true response kernel $\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}$. This last form is the one used to build the densities, potentials, and Hamiltonian operators (both zero-order and first-order).

4.4.3 Variational second-order energy

The most difficult part in converting the variational second-order energy (Eq. 4.2.22) to NGWF format is dealing with the terms quadratic in first-order wavefunctions. For the variational form we will only need the diagonal perturbations $\frac{d^2}{d\lambda_{-\mathbf{q}}d\lambda_{\mathbf{q}}}E_{\mathrm{TOT}}$, which greatly simplifies our task. We thus note the following:

$$\mathrm{Tr}\left\{\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\mathbf{S}\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi\ \dagger}\mathbf{H}^\theta_{\mathbf{q},\mathbf{q}}\right\} = \left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\alpha\gamma}S_{\gamma\delta}\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \delta\beta}\left(H^\theta_{\mathbf{q},\mathbf{q}}\right)_{\beta\alpha} =$$

$$\sum_i^{N/2}\sum_j^{N/2}\underbrace{\langle\widetilde{\psi}_i|\,\hat{S}\,|\widetilde{\psi}_j\rangle}_{\delta_i j}\langle\overline{\widetilde{\psi}_j^{1\lambda_{\mathbf{q}}}}|\,\widetilde{H}_{\mathbf{q},\mathbf{q}}\,|\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}\rangle = \sum_i^{N/2}\langle\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}|\,\widetilde{H}_{\mathbf{q},\mathbf{q}}\,|\overline{\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}}\rangle \tag{4.4.17}$$

and

$$\mathrm{Tr}\Big\{\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\mathbf{H}\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi\ \dagger}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\Big\} = \Big(P^{1\lambda_\mathbf{q}\ \theta\phi}\Big)^{\alpha\gamma} H_{\gamma\delta}\Big(P^{1\lambda_\mathbf{q}\ \theta\phi}\Big)^{\dagger\ \delta\beta}\Big(S^{\theta}_{\mathbf{q},\mathbf{q}}\Big)_{\beta\alpha} = \sum_{i}^{N/2}\sum_{j}^{N/2}\underbrace{\langle\widetilde{\psi}_i|\,\widetilde{H}\,|\widetilde{\psi}_j\rangle}_{\epsilon_i\delta_i j}\langle\overline{\widetilde{\psi}^{1\lambda_\mathbf{q}}_j}|\,\widetilde{S}_{\mathbf{q},\mathbf{q}}\,|\overline{\widetilde{\psi}^{1\lambda_\mathbf{q}}_i}\rangle = \sum_{i}^{N/2}\epsilon_i\,\langle\overline{\widetilde{\psi}^{1\lambda_\mathbf{q}}_i}|\,\hat{S}_{\mathbf{q},\mathbf{q}}\,|\overline{\widetilde{\psi}^{1\lambda_\mathbf{q}}_i}\rangle \tag{4.4.18}$$

Since $\widetilde{H}_{\mathbf{q},\mathbf{q}}$ and $\hat{S}_{\mathbf{q},\mathbf{q}}$ are hermitian, ϵ_i is real, and $|\overline{\widetilde{\psi}^{1\lambda_\mathbf{q}}_i}\rangle^{\dagger} = \langle\overline{\widetilde{\psi}^{1\lambda_\mathbf{q}}_i}|$ (where † implies the conjugate transpose), it is clear that the terms in Eqs. 4.4.17 and 4.4.18 are real. These expressions match the quadratic terms in the variational second-order energy, apart from the spin factors. Thus, remembering that the first-order Lagrange multipliers in Eq. 4.2.22 disappear in the parallel-transport gauge, one obtains the NGWF-form of the variational second-order energy ($s = 2$ is the spin factor for spin-unpolarised systems):

$$\begin{aligned}
&\frac{d^2\ E_{\mathrm{Var}}}{d\lambda_{-\mathbf{q}}d\lambda_\mathbf{q}} = s\mathrm{Tr}\Big\{\mathbf{K}\Big[\mathbf{A}^{\lambda_{-\mathbf{q}}\lambda_\mathbf{q}} - \mathbf{H}\mathbf{S}^{-1}\mathbf{S}^{2\lambda_{-\mathbf{q}}\lambda_\mathbf{q}}\Big]\Big\} + 2s\mathrm{Tr}\Big\{\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\mathbf{S}\Big(\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\Big)^{\dagger}\mathbf{H}^{\theta}_{\mathbf{q},\mathbf{q}}\Big\} - \\
&2s\mathrm{Tr}\Big\{\Big(\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\mathbf{H}\Big(\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\Big)^{\dagger}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\Big)\Big\} + 2s\mathrm{Tr}\Big\{\big(\mathbf{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\big)^{\phi\theta}\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\Big\} + 2s\mathrm{Tr}\Big\{\Big(\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\Big)^{\dagger}\big(\mathbf{B}^{\lambda_\mathbf{q}}_{\mathbf{q},\mathbf{0}}\big)^{\theta\phi}\Big\} - \\
&2s\mathrm{Tr}\Big\{\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\mathbf{H}\mathbf{S}^{-1}\big(\mathbf{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\big)^{\phi\theta}\Big\} - 2s\mathrm{Tr}\Big\{\big(\mathbf{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\big)^{\phi\theta\ \dagger}\,\mathbf{S}^{-1}\mathbf{H}\Big(\mathbf{P}^{1\lambda_\mathbf{q}\ \theta\phi}\Big)^{\dagger}\Big\} + \\
&\int\overline{v^{1\lambda_{-\mathbf{q}}}_{\mathrm{XC}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\overline{\widetilde{\rho}^{1\lambda_\mathbf{q}}_c}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \int v_{XC}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\frac{\partial^2\widetilde{\rho}_c}{\partial\lambda_{-\mathbf{q}}\partial\lambda_\mathbf{q}}(\mathbf{r})\ \mathrm{d}\mathbf{r} + \frac{\partial^2\ U}{\partial\lambda_{-\mathbf{q}}\lambda_\mathbf{q}} + \\
&2s\mathrm{Tr}\Big\{\frac{\partial\overrightarrow{E}}{\partial\lambda_\mathbf{0}}\cdot\Big[\overrightarrow{\mathbf{D}}^{\theta\phi}\Big(P^{1\lambda_0\ \theta\phi}\Big)^{\dagger} + \overrightarrow{\mathbf{D}}^{\phi\theta}P^{1\lambda_0\ \theta\phi}\Big]\Big\}\,\delta_{\mathbf{q},\mathbf{0}}
\end{aligned} \tag{4.4.19}$$

We remind the reader that $\hat{A}^{\lambda_{-\mathbf{q}}\lambda_\mathbf{q}}$ is defined in Eq. 4.2.28, $\hat{B}^{1\lambda_\mathbf{q}}_{\mathbf{q},\mathbf{0}}$ in Eq. 4.2.29, and $\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}$ in Eq. 4.2.46. In the previous equation, terms in red have no dependence on the first-order NGWFs.

4.4.4 Sternheimer Equation

We obtain the Sternheimer equation (Eq. 4.2.38) in the NGWF framework by minimising the variational second-order energy (Eq. 4.4.19) with respect to the auxiliary response kernel $\underline{P}^{1\lambda_\mathbf{q}\theta\phi}$. Therefore, one first needs to replace the response kernel $P^{1\lambda_\mathbf{q}\theta\phi}$ from Eq. 4.4.19 with the gauge constraint expression that contains the auxiliary version (Eq. 4.4.16). Furthermore, all first-order density terms depend on the adjoint $\underline{P}^{1\lambda_\mathbf{q}\theta\phi\ \dagger}$

(Eqs. 4.2.8 and 4.4.14), and it is standard practice to consider $\underline{P}^{1\lambda_{\mathbf{q}}\theta\phi}$ as independent of $\underline{P}^{1\lambda_{\mathbf{q}}\theta\phi\,\dagger}$ in variational problems. Thus, the minimisation of the variational second-order energy

$$\frac{\delta}{\delta\left(\underline{P}^{1\lambda_{\mathbf{q}}\theta\phi}\right)^{\alpha\beta}}\frac{d^2\ E_{\mathrm{Var}}}{d\lambda_{-\mathbf{q}}d\lambda_{\mathbf{q}}}=0 \tag{4.4.20}$$

yields the sought-after Sternheimer equation in matrix format:

$$\begin{aligned}&\mathbf{SKS}\left(\underline{\mathbf{P}}^{1\lambda_{\mathbf{q}}\theta\phi}\right)^{\dagger}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{H}^{\theta}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}-\mathbf{SKH}\left(\mathbf{P}^{1\lambda_{\mathbf{q}}\theta\phi}\right)^{\dagger}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\underbrace{\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}}_{\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}=\\&-\mathbf{SKH}^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}+\mathbf{HKS}^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}-\mathbf{SK}\left(\delta_{\mathbf{q},\mathbf{0}}\frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\mathbf{D}^{\phi\theta}\right)\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\end{aligned} \tag{4.4.21}$$

Within the minimisation, we have also used $\mathbf{K}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}=\mathbf{K}_{\mathbf{q},\mathbf{q}}\mathbf{H}^{\phi\theta}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}=\mathbf{0}$, due to the orthogonality between the valence and "out-of-valence" manifolds at wave-vector $\mathbf{q}$.

We can apply $\mathbf{S}^{-1}$ from the left without changing the result, but a right-hand side simplification requires more thought. If we were to apply the inverse overlap in the joint NGWF-basis $\mathbf{S}^{\theta\ -1}_{\mathbf{q},\mathbf{q}}$ from the right, we would essentially mix elements from the zero-order and first-order NGWF vector-spaces, which has proven to be devastating on the optimisation of first-order NGWFs $\delta\phi$. Thus, one needs to project in disjoint NGWF manifolds by multiplying with a block-diagonal matrix:

$$\begin{bmatrix}\left(S^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1} & 0\\ 0 & S^{-1}_{\mathbf{q},\mathbf{q}}\end{bmatrix} \tag{4.4.22}$$

By applying the aformentioned matrix to the right, as well as a conjugate transpose operation on the entire object, we end up with the final form of the Sternheimer equation, which takes a generalised Sylvester form:

$$\mathbf{A}^{\theta}\mathbf{X}^{\theta\phi}\mathbf{B}+\mathbf{C}^{\theta}\mathbf{X}^{\theta\phi}\mathbf{D}=\underbrace{\mathbf{E}^{\theta\phi}_{\mathrm{fixed}}+\mathbf{E}^{\theta\phi}_{\mathrm{variable}}}_{\mathbf{E}^{\theta\phi}}\ , \tag{4.4.23}$$

with

$$\begin{aligned}&\mathbf{X}^{\theta\phi}=\underline{\mathbf{P}}^{1\lambda_{-\mathbf{q}}\ \theta\phi}\qquad\mathbf{A}^{\theta}=\begin{bmatrix}\left(\mathbf{S}^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1}\mathbf{S}^{\delta\phi\theta}_{\mathbf{q},\mathbf{q}}\\ \mathbf{S}^{-1}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}}\end{bmatrix}\ \mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{H}^{\theta}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\qquad\mathbf{B}=\mathbf{SK}\\&\mathbf{C}^{\theta}=-\begin{bmatrix}\left(\mathbf{S}^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1}\mathbf{S}^{\delta\phi\theta}_{\mathbf{q},\mathbf{q}}\\ \mathbf{S}^{-1}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}}\end{bmatrix}\left[\left(\mathbf{S}^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1}\mathbf{S}^{\delta\phi\theta}_{\mathbf{q},\mathbf{q}}\ ;\ \mathbf{S}^{-1}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}}\right]\mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\mathbf{S}^{\theta}_{\mathbf{q},\mathbf{q}}\qquad\mathbf{D}=\mathbf{SKHK}\end{aligned} \tag{4.4.24}$$

$$\mathbf{E}^{\theta\phi}_{\text{fixed}} = \begin{bmatrix} \left(\mathbf{S}^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1}\mathbf{S}^{\delta\phi\theta}_{\mathbf{q},\mathbf{q}} \\ \mathbf{S}^{-1}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}} \end{bmatrix} \mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\left(\mathbf{S}^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}\mathbf{S}^{-1}\mathbf{H}\mathbf{K} - \begin{bmatrix} \left(\mathbf{S}^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1}\mathbf{S}^{\delta\phi\theta}_{\mathbf{q},\mathbf{q}} \\ \mathbf{S}^{-1}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}} \end{bmatrix} \mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\left(\frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\overrightarrow{\mathcal{D}}^{\theta\phi}\right)\mathbf{K}\delta_{\mathbf{q},\mathbf{0}}$$

$$\mathbf{E}^{\theta\phi}_{\text{variable}} = -\begin{bmatrix} \left(\mathbf{S}^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)^{-1}\mathbf{S}^{\delta\phi\theta}_{\mathbf{q},\mathbf{q}} \\ \mathbf{S}^{-1}_{\mathbf{q},\mathbf{q}}\mathbf{S}^{\phi\theta}_{\mathbf{q},\mathbf{q}} \end{bmatrix} \mathbf{K}^{\theta}_{c\ \mathbf{q},\mathbf{q}}\left(\mathbf{H}^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}\mathbf{K}$$

It is thus clear that we optimise the auxiliary response kernel $\underline{\mathbf{P}}^{1\lambda_{-\mathbf{q}}\ \theta\phi}$. Matrix $\mathbf{E}^{\theta\phi}_{\text{fixed}}$ does not require self-consistency, but $\mathbf{E}^{\theta\phi}_{\text{variable}}$ depends self-consistently on the response kernel $\mathbf{P}^{1\lambda_{-\mathbf{q}}\ \theta\phi}$ (obtained from $\underline{\mathbf{P}}^{1\lambda_{-\mathbf{q}}\ \theta\phi}$) through $\mathbf{H}^{1\lambda_{-\mathbf{q}}\ \phi\theta}_{\mathbf{0},\mathbf{q}}$. Note that $\overrightarrow{E}$ is the external electric field (if present at all), while $\overrightarrow{\mathcal{D}}^{\theta\phi}$ (Eq. 4.5.14) is the vector of dipole matrices, in joint-valence NGWF representation.

We stress that Eq. 4.4.23 is highly non-standard, and is not encountered in traditional-DFT formalisms for linear-response phonons. The reason for this unique formulation comes from the need to eliminate explicit eigenvalues in a linear-scaling DFT formalism; in turn, this determines the appearance of matrices **B** and **D** (Eq. 4.4.24). If the latter were both the identity matrix, Eq. 4.4.23 could have been solved through a standard conjugate-gradient approach. However, it turns out that we need to employ a quasi-Newton optimisation approach to solve for $\mathbf{X}^{\theta\phi}$, as explained in Section 4.6.

4.4.5 Non-variational second-order energy

By transforming Eqs. 4.2.42 and 4.2.45 into a NGWF-basis formulation, one easily obtains:

$$\frac{d^2\ E_{\text{Non-var}}}{d\lambda_{-\mathbf{q}}d\epsilon_{\mathbf{q}}} = s\text{Tr}\left\{\mathbf{K}\left[\mathbf{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}} - \mathbf{H}\mathbf{S}^{-1}\mathbf{S}^{2\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}}\right]\right\} + 2s\text{Tr}\left\{\left(\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}(\mathbf{B}^{\epsilon_{\mathbf{q}}})^{\theta\phi}_{\mathbf{q},\mathbf{0}}\right\} -$$
$$2s\text{Tr}\left\{(\mathbf{S}^{1\epsilon_{\mathbf{q}}})^{\theta\phi}_{\mathbf{q},\mathbf{0}}\ \mathbf{S}^{-1}\mathbf{H}\left(\mathbf{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right\} + \int \overline{v^{1\lambda_{-\mathbf{q}}}_{\text{XC}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\overline{\widetilde{\rho}^{1\epsilon_{\mathbf{q}}}_c}(\mathbf{r})\ \text{d}\mathbf{r} + \tag{4.4.25}$$
$$\int v_{XC}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\frac{\partial^2\widetilde{\rho}_c}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}(\mathbf{r})\ \text{d}\mathbf{r} + \frac{\partial^2\ U}{\partial\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}} + 2s\text{Tr}\left\{\frac{\partial\overrightarrow{E}}{\partial\epsilon_{\mathbf{0}}}\cdot\overrightarrow{\mathbf{D}}^{\theta\phi}\left(\mathbf{P}^{1\lambda_{\mathbf{0}}\ \theta\phi}\right)^{\dagger}\right\}\delta_{\mathbf{q},\mathbf{0}} +$$
$$\frac{1}{2}s\text{Tr}\left\{\mathbf{K}\mathbf{H}\mathbf{S}^{-1}\left(\mathbf{S}^{1\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\right)^{\theta\phi\ \dagger}\mathbf{K}^{\theta}_{v\ \mathbf{q},\mathbf{q}}\left(\mathbf{S}^{1\epsilon_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\right)^{\theta\phi}\right\} + \frac{1}{2}s\text{Tr}\left\{\mathbf{K}\left(\mathbf{S}^{1\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\right)^{\theta\phi\ \dagger}\mathbf{S}^{\theta\ -1}_{\mathbf{q},\mathbf{q}}\mathbf{H}_{\mathbf{q},\mathbf{q}}\mathbf{K}^{\theta}_{v\ \mathbf{q},\mathbf{q}}\left(\mathbf{S}^{1\epsilon_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\right)^{\theta\phi}\right\} -$$
$$s\text{Tr}\left\{\mathbf{K}\left(\mathbf{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\phi\theta}\mathbf{K}^{\theta}_{v\ \mathbf{q},\mathbf{q}}\left(\mathbf{S}^{1\epsilon_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\right)^{\theta\phi}\right\}$$

4.4.6 First-order NGWF gradient

Apart from solving for the response kernel, we also need to optimise the first-order NGWFs $\delta\phi$, as one needs both components to obtain the first-order wavefunction. The dependence of the variational second-order energy (Eq. 4.4.19) on $\delta\phi$ comes in two flavours: explicit (through matrices in the joint set θ representation) and implicit (through first-order densities). Thus, for clarity, we calculate the different gradient components separately. Varying the energy with respect to $\langle\theta|$ implies

$$|\big(g^{\theta}\big)^{\eta}\rangle = \frac{\delta}{\delta\langle\theta_\eta|}\left[\frac{d^2E_{\mathrm{TOT}}}{d\lambda_{-\mathbf{q}}d\lambda_{\mathbf{q}}}\right] = |\big(g^{\theta}_{\mathrm{IMP}}\big)^{\eta}\rangle + |\big(g^{\theta}_{\mathrm{EXP}}\big)^{\eta}\rangle \tag{4.4.26}$$

For the remainder of this section, when the superscript "$1\lambda_{-\mathbf{q}}$" appears above a position-dependent quantity (such as a density or potential), it implies that perturbation $\frac{d}{d\lambda_{-\mathbf{q}}}$ has been applied. When a matrix element has that superscript, it only means that the operator associated to the matrix was perturbed, i.e. $\left(H^{1\lambda_{-\mathbf{q}}\ \theta\phi}_{\mathbf{0},\mathbf{q}}\right)_{\alpha\beta} = \langle\theta_\alpha|\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}|\phi_\beta\rangle$.

Implicit terms

$$\begin{aligned}
|\big(g^{\theta}_{\mathrm{IMP}}\big)^{\eta}\rangle =& sK^{\beta\alpha}\frac{\delta A^{\lambda_{-\mathbf{q}}\lambda_{\mathbf{q}}}_{\alpha\beta}}{\delta\langle\theta_\eta|} + 2s\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\langle\theta_\beta|\frac{\delta\ \hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}}{\delta\langle\theta_\eta|}|\phi_\alpha\rangle + \\
& 2s\langle\phi_\alpha|\frac{\delta H^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}}{\delta\langle\theta_\eta|}|\theta_\beta\rangle\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\beta\alpha} + 2s\frac{\delta\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\langle\theta_\beta|}{\delta\langle\theta_\eta|}\hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}|\phi_\alpha\rangle + \\
& 2s\langle\phi_\alpha|\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\frac{\delta|\theta_\beta\rangle\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\beta\alpha}}{\delta\langle\theta_\eta|} + \int\frac{\delta\overline{v^{1\lambda_{-\mathbf{q}}}_{\mathrm{XC}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\langle\theta_\eta|}\overline{\widetilde{\rho}^{1\lambda_{\mathbf{q}}}_c}(\mathbf{r}')\ \mathrm{d}\mathbf{r}'\ ,
\end{aligned} \tag{4.4.27}$$

where we have implicitly used the variation chain rule:

$$\frac{\delta}{\delta\langle\theta_\eta|}\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi\ \dagger}B^{\lambda_{\mathbf{q}}\ \theta\phi}_{\mathbf{q},\mathbf{0}}\big)^{\alpha}_{\ \alpha} = \frac{\delta\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\langle\theta_\beta|}{\delta\langle\theta_\eta|}\hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}|\phi_\alpha\rangle + \big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\langle\theta_\beta|\frac{\delta\ \hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}}{\delta\langle\theta_\eta|}|\phi_\alpha\rangle\ , \tag{4.4.28}$$

but the variation of contra-variant matrix elements, such as $\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\alpha\beta}$ is more complex and will be discussed later.

By looking at the definitions of the $A^{\lambda_{-\mathbf{q}}\lambda_{\mathbf{q}}}$ and $B^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}$ operators in the main text (Eqs. 4.2.28 and 4.2.29 where ϵ was replaced with λ), as well as the form of $\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}$ in Eq. 4.2.46,

one can expand this gradient component as:

$$|\big(g_{\mathrm{IMP}}^{\delta\phi}\big)^{\eta}\rangle = s\sum_{a,b} K^{\alpha\beta}\,\langle\phi_\beta|\widetilde{p}^{a}\rangle\,\langle\widetilde{p}^{b}|\phi_\alpha\rangle \int \frac{\delta\overline{\widetilde{v}_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')}{\delta\,\langle\theta_\eta|}\overline{\hat{Q}_{ab}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')\;\mathrm{d}\mathbf{r}' + s\sum_{a,b} K^{\alpha\beta}\Big[\,\langle\phi_\beta|\widetilde{p}^{a\ 1\lambda_{\mathbf{q}}}\rangle\,\langle\widetilde{p}^{b}|\phi_\alpha\rangle +$$
$$\langle\phi_\beta|\widetilde{p}^{b\ 1\lambda_{\mathbf{q}}}\rangle\,\langle\widetilde{p}^{a}|\phi_\alpha\rangle\,\Big]\frac{\delta D_{ab}^{1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|} + 2s\,\langle\phi_\alpha|\frac{\delta\overline{\widetilde{v}_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')}{\delta\,\langle\theta_\eta|}|\theta_\beta\rangle\,\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\beta\alpha} + \tag{4.4.29}$$
$$2s\sum_{a,b}\langle\phi_\alpha|\widetilde{p}^{a}\rangle\,\langle\widetilde{p}^{b}_{\mathbf{q}}|\theta_\beta\rangle\,\frac{\delta D_{ab}^{1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|}\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\beta\alpha} + \int\frac{\delta\overline{v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\,\langle\theta_\eta|}\overline{\widetilde{\rho}_c^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')\;\mathrm{d}\mathbf{r}' +$$
$$2s\frac{\delta\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\,\langle\theta_\beta|}{\delta\,\langle\theta_\eta|}\;B^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\,|\phi_\alpha\rangle + 2s\,\langle\phi_\alpha|\,H^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\;\underbrace{\frac{\delta\,|\theta_\beta\rangle\,\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\beta\alpha}}{\delta\,\langle\theta_\eta|}}_{=0}$$

By remembering the definitions of the zero-order and first-order projector density-kernel $\rho^{ab\ 1\lambda_{-\mathbf{q}}}$ (Eq. 4.2.10), as well as the first-order pseudo-density (Eq. 4.2.7), we can coalesce some terms into a simpler expression:

$$|\big(g_{\mathrm{IMP}}^{\delta\phi}\big)^{\eta}\rangle = \int\frac{\delta\overline{\widetilde{v}_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')}{\delta\,\langle\theta_\eta|}\Big[\overline{\widetilde{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \sum_{a,b}\rho^{ab}\overline{\hat{Q}_{ab}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')\Big]\;\mathrm{d}\mathbf{r}' + \int\frac{\delta\overline{v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\,\langle\theta_\eta|}\overline{\widetilde{\rho}_c^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')\;\mathrm{d}\mathbf{r}' +$$
$$\sum_{a,b}\rho^{ab\ 1\lambda_{\mathbf{q}}}\frac{\delta D_{ab}^{1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|} + 2s\frac{\delta\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\,\langle\theta_\beta|}{\delta\,\langle\theta_\eta|}\;\hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\,|\phi_\alpha\rangle \tag{4.4.30}$$

We can make further simplifications by observing that the first-order pseudo-potential $\overline{\widetilde{v}_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}_{Z_c}]$ has no dependence on θ, implying:

$$\frac{\delta\overline{\widetilde{v}_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')}{\delta\,\langle\theta_\eta|} = \frac{\delta\overline{\widetilde{v}_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}')}{\delta\,\langle\theta_\eta|} + \frac{\delta\overline{\widetilde{v}_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\,\langle\theta_\eta|} \tag{4.4.31}$$

Lastly, we can expand $D_{ab}^{1\lambda_{-\mathbf{q}}}$ as described in Sec. 4.2.5:

$$\frac{\delta}{\delta\,\langle\theta_\eta|}D_{ab}^{1\lambda_{-\mathbf{q}}} = \frac{\delta}{\delta\,\langle\theta_\eta|}\Big\{\int\Big[\overline{\widetilde{v}_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')\hat{Q}_{ab,\mathbf{q}}(\mathbf{r}') + \widetilde{v}_{\mathrm{eff}}(\mathbf{r}')\overline{\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')\Big]\;\mathrm{d}\mathbf{r}' + D_{\mathrm{H}\ ab}^{1\lambda_{-\mathbf{q}}} + D_{\mathrm{XC}\ ab}^{1\lambda_{-\mathbf{q}}}\Big\} =$$
$$\int\frac{\delta\overline{\widetilde{v}_{\mathrm{eff}}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')}{\delta\,\langle\theta_\eta|}\hat{Q}_{ab,\mathbf{q}}(\mathbf{r}')\;\mathrm{d}\mathbf{r}' + \frac{\delta}{\delta\,\langle\theta_\eta|}\Big[D_{\mathrm{H}\ ab}^{1\lambda_{-\mathbf{q}}} + D_{\mathrm{XC}\ ab}^{1\lambda_{-\mathbf{q}}}\Big] \tag{4.4.32}$$

However, by looking at the expression of $D_{\mathrm{H}\ ab}^{1\lambda_{-\mathbf{q}}}$ and $D_{\mathrm{XC}\ ab}^{1\lambda_{-\mathbf{q}}}$ in Sec. 4.2.5, it is clear that their sum can be expressed as a tensor contraction $\sum_{i,j}\rho^{ij\ 1\lambda_{-\mathbf{q}}}\,t_{ijab}$, meaning that

$$\sum_{a,b}\rho^{ab\ 1\lambda_{\mathbf{q}}}\frac{\delta}{\delta\,\langle\theta_\eta|}\Big[D_{\mathrm{H}\ ab}^{1\lambda_{-\mathbf{q}}} + D_{\mathrm{XC}\ ab}^{1\lambda_{-\mathbf{q}}}\Big] = \sum_{a,b,i,j}\rho^{ab\ 1\lambda_{\mathbf{q}}}\frac{\delta\rho^{ij\ 1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|}t_{ijab} = \tag{4.4.33}$$
$$\sum_{i,j}\frac{\delta\rho^{ij\ 1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|}\Big[D_{\mathrm{H}\ ij}^{1\lambda_{\mathbf{q}}} + D_{\mathrm{XC}\ ij}^{1\lambda_{\mathbf{q}}}\Big] = \sum_{a,b}\frac{\delta\rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|}\Big[D_{\mathrm{H}\ ab}^{1\lambda_{\mathbf{q}}} + D_{\mathrm{XC}\ ab}^{1\lambda_{\mathbf{q}}}\Big]$$

By applying Eqs. 4.4.31, 4.4.32, 4.4.33 to the implicit NGWF gradient (Eq. 4.4.30) we arrive at:

$$|\big(g^{\delta\phi}_{\mathrm{IMP}}\big)^{\eta}\rangle = \int \frac{\delta\overline{\widetilde{v}_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}')}{\delta\,\langle\theta_\eta|}\Big[\overline{\widetilde{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \underbrace{\sum_{a,b}\rho^{ab}\overline{\hat{Q}_{ab}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \sum_{a,b}\rho^{ab\ 1\lambda_{\mathbf{q}}}\hat{Q}_{ab,\mathbf{q}}(\mathbf{r}')}_{\overline{\hat{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')}\Big]\,\mathrm{d}\mathbf{r}' +$$

$$\int \frac{\delta\overline{\widetilde{v}_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\,\langle\theta_\eta|}\Big[\overline{\widetilde{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \underbrace{\sum_{a,b}\rho^{ab}\overline{\hat{Q}_{ab}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \sum_{a,b}\rho^{ab\ 1\lambda_{\mathbf{q}}}\hat{Q}_{ab,\mathbf{q}}(\mathbf{r}')}_{\overline{\hat{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')} + \overline{\widetilde{\rho}_c^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')\Big]\,\mathrm{d}\mathbf{r}' +$$

$$\sum_{a,b}\frac{\delta\rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta\,\langle\theta_\eta|}\Big[D^{1\lambda_{\mathbf{q}}}_{\mathrm{H}\ ab} + D^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}\ ab}\Big] + 2s\frac{\delta\big(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\big)^{\dagger\ \alpha\beta}\ \langle\theta_\beta|}{\delta\,\langle\theta_\eta|}\ B^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}\,|\phi_\alpha\rangle \tag{4.4.34}$$

Also, we can expand these first-order Hartree and XC potentials into their kernel expressions (note that the presence or absence of phase factorisation is highly relevant):

$$\begin{aligned}
&\overline{\widetilde{v}_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}') = e^{i\mathbf{q}\cdot\mathbf{r}'}\ \widetilde{v}_{\mathrm{H}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}') = \\
&\qquad e^{i\mathbf{q}\cdot\mathbf{r}'}\int f_{\mathrm{H}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}',\mathbf{r}'')\big[\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'') + \hat{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'')\big]\,\mathrm{d}\mathbf{r}'' \\
&\overline{\widetilde{v}_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}') = e^{i\mathbf{q}\cdot\mathbf{r}'}\ \widetilde{v}_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}') = \\
&\qquad e^{i\mathbf{q}\cdot\mathbf{r}'}\int f_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}',\mathbf{r}'')\big[\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'') + \hat{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'') + \widetilde{\rho}_c^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'')\big]\,\mathrm{d}\mathbf{r}''\ ,
\end{aligned} \tag{4.4.35}$$

where $f_{\mathrm{H/XC}}$ is simply the functional derivative of $\widetilde{v}_{\mathrm{H/XC}}$. Therefore, Eq. 4.4.35 allows us to rewrite some terms from the implicit gradient as:

$$\begin{aligned}
&\int \frac{\delta\overline{\widetilde{v}_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\,\langle\theta_\eta|}\Big[\overline{\widetilde{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \overline{\hat{\rho}^{1\lambda_{\mathbf{q}}}}(\mathbf{r}') + \overline{\widetilde{\rho}_c^{1\lambda_{\mathbf{q}}}}(\mathbf{r}')\Big]\,\mathrm{d}\mathbf{r}' = \\
&\int \frac{\delta\widetilde{v}_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')}{\delta\,\langle\theta_\eta|}\Big[\widetilde{\rho}^{1\lambda_{\mathbf{q}}}(\mathbf{r}') + \hat{\rho}^{1\lambda_{\mathbf{q}}}(\mathbf{r}') + \widetilde{\rho}_c^{1\lambda_{\mathbf{q}}}(\mathbf{r}')\Big]\,\mathrm{d}\mathbf{r}' = \\
&\int\int f_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}',\mathbf{r}'')\frac{\delta[\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}+\hat{\rho}^{1\lambda_{-\mathbf{q}}}+\widetilde{\rho}_c^{1\lambda_{-\mathbf{q}}}](\mathbf{r}'')}{\delta\,\langle\theta_\eta|}\Big[\widetilde{\rho}^{1\lambda_{\mathbf{q}}}+\hat{\rho}^{1\lambda_{\mathbf{q}}}+\widetilde{\rho}_c^{1\lambda_{\mathbf{q}}}\Big](\mathbf{r}')\,\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}'' = \\
&\int \widetilde{v}_{\mathrm{XC}}^{1\lambda_{\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}'')\frac{\delta[\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'')+\hat{\rho}^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'')+\widetilde{\rho}_c^{1\lambda_{-\mathbf{q}}}(\mathbf{r}'')]}{\delta\,\langle\theta_\eta|}\,\mathrm{d}\mathbf{r}'' = \\
&\int \overline{\widetilde{v}_{\mathrm{XC}}^{1\lambda_{\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}')\frac{\delta[\overline{\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')+\overline{\hat{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')]}{\delta\,\langle\theta_\eta|}\,\mathrm{d}\mathbf{r}'
\end{aligned} \tag{4.4.36}$$

where we have implicitly used that $\frac{\delta}{\delta\langle\theta_\eta|}\widetilde{\rho}_c^{1\lambda_{-\mathbf{q}}} = 0$. An identical procedure can be applied to the first-order Hartree potential term in Eq. 4.4.34. Thus, we end up with:

$$|\big(g^{\delta\phi}_{\mathrm{IMP}}\big)^{\eta}\rangle = \int \Big[\overline{\widetilde{v}_{\mathrm{XC}}^{1\lambda_{\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r}') + \overline{\widetilde{v}_{\mathrm{H}}^{1\lambda_{\mathbf{q}}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r}')\Big]\frac{\delta[\overline{\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')+\overline{\hat{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}')]}{\delta\,\langle\theta_\eta|}\,\mathrm{d}\mathbf{r}' +$$

$$\sum_{a,b} \frac{\delta \rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta \langle \theta_\eta |} \left[D^{1\lambda_{\mathbf{q}}}_{\mathrm{H}\ ab} + D^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}\ ab} \right] + 2s \frac{\delta \left(P^{1\lambda_{\mathbf{q}}\ \theta\phi} \right)^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} \ B^{\lambda_{\mathbf{q}}\ \theta\phi}_{\mathbf{q},\mathbf{0}} |\phi_\alpha\rangle \quad (4.4.37)$$

Moreover, by expanding the variation of the first-order densities as

$$\frac{\delta \overline{\widetilde{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}'')}{\delta \langle \theta_\eta |} = 2s \langle \mathbf{r}'' | \phi_\alpha \rangle \frac{\delta \left(P^{1\lambda_{\mathbf{q}}\ \theta\phi} \right)^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} | \mathbf{r}'' \rangle \quad (4.4.38)$$

and

$$\frac{\delta \overline{\hat{\rho}^{1\lambda_{-\mathbf{q}}}}(\mathbf{r}'')}{\delta \langle \theta_\eta |} = \sum_{a,b} \left[\frac{\delta \rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta \langle \theta_\eta |} \hat{Q}_{ab,-\mathbf{q}}(\mathbf{r}'') + \rho^{ab} \underbrace{\frac{\delta \overline{\hat{Q}^{1\lambda_{-\mathbf{q}}}_{ab}}(\mathbf{r}'')}{\delta \langle \theta_\eta |}}_{0} \right] = \sum_{a,b} \frac{\delta \rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta \langle \theta_\eta |} \hat{Q}_{ab,-\mathbf{q}}(\mathbf{r}'') \ , \quad (4.4.39)$$

one obtains

$$\left| \left(g^{\delta\phi}_{\mathrm{IMP}} \right)^{\eta} \right\rangle = 2s \frac{\delta (P^{1\lambda_{\mathbf{q}}\ \theta\phi})^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} \left[\overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{H}}} + \overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}}} \right] |\phi_\alpha\rangle + 2s \frac{\delta \left(P^{1\lambda_{\mathbf{q}}\ \theta\phi} \right)^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} \ B^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}} |\phi_\alpha\rangle +$$
$$\sum_{a,b} \frac{\delta \rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta \langle \theta_\eta |} \left[\int \left(\overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{H}}}(\mathbf{r}') + \overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}}}(\mathbf{r}') \right) \hat{Q}_{ab,-\mathbf{q}}(\mathbf{r}') \ \mathrm{d}\mathbf{r}' + D^{1\lambda_{\mathbf{q}}}_{\mathrm{H}\ ab} + D^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}\ ab} \right] \quad (4.4.40)$$

By expressing the variation of the first-order projector density kernel (Eq. 4.2.11):

$$\frac{\delta \rho^{ab\ 1\lambda_{-\mathbf{q}}}}{\delta \langle \theta_\eta |} = s \frac{\delta \left(P^{1\lambda_{\mathbf{q}}\ \theta\phi} \right)^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} \left[|\widetilde{p}^b_{\mathbf{q}}\rangle \langle \widetilde{p}^a | + |\widetilde{p}^a_{\mathbf{q}}\rangle \langle \widetilde{p}^b | \right] |\phi_\alpha\rangle \quad (4.4.41)$$

one finally obtains:

$$\left| \left(g^{\delta\phi}_{\mathrm{IMP}} \right)^{\eta} \right\rangle = s \frac{\delta (P^{1\lambda_{\mathbf{q}}\ \theta\phi})^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} \Bigg\{ 2\overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{H}}} + 2\overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}}} + 2\hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}} + \sum_{a,b} \left(|\widetilde{p}^b_{\mathbf{q}}\rangle \langle \widetilde{p}^a | + |\widetilde{p}^a_{\mathbf{q}}\rangle \langle \widetilde{p}^b | \right)$$
$$\left[\int \left(\overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{H}}}(\mathbf{r}') + \overline{\widetilde{v}^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}}}(\mathbf{r}') \right) \hat{Q}_{ab,-\mathbf{q}}(\mathbf{r}') \ \mathrm{d}\mathbf{r}' + D^{1\lambda_{\mathbf{q}}}_{\mathrm{H}\ ab} + D^{1\lambda_{\mathbf{q}}}_{\mathrm{XC}\ ab} \right] \Bigg\} |\phi_\alpha\rangle \quad (4.4.42)$$

However, a keen observer can see that the operators marked by red are actually identical to $2\left(\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}$ (see the form of $\hat{B}^{\lambda_{\mathbf{q}}}_{\mathbf{q},\mathbf{0}}$ in Eq. 4.2.29), leaving us with a compact and simple form for the implicit NGWF gradient terms:

$$\left| \left(g^{\delta\phi}_{\mathrm{IMP}} \right)^{\eta} \right\rangle = 2s \frac{\delta (P^{1\lambda_{\mathbf{q}}\ \theta\phi})^{\dagger\ \alpha\beta} \langle \theta_\beta |}{\delta \langle \theta_\eta |} \left(\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}} \right)^{\dagger} |\phi_\alpha\rangle \quad (4.4.43)$$

In an almost identical manner, one can prove that the complex conjugate of the implicit gradient (obtained through $\frac{\delta}{\delta |\theta_\eta\rangle}$) is:

$$\left\langle \left(g^{\delta\phi}_{\mathrm{IMP}} \right)^{\eta} \right| = 2s \langle \phi_\alpha | \widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}} \frac{\delta |\theta_\beta\rangle \ (P^{1\lambda_{\mathbf{q}}\ \theta\phi})^{\beta\alpha}}{\delta |\theta_\eta\rangle} \quad (4.4.44)$$

Explicit terms

The remaining term of the gradient, containing only explicit dependencies of the joint NGWFs θ, can be easily obtained as:

$$|\left(g^{\delta\phi}_{\mathrm{EXP}}\right)^{\eta}\rangle = 2s\frac{\delta\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \alpha\beta}\langle\theta_\beta|}{\delta\langle\theta_\eta|}\Bigg[\widetilde{H}_{\mathbf{q},\mathbf{q}}|\theta_\gamma\rangle\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}S\right)^{\gamma}_{\ \alpha} - \hat{S}_{\mathbf{q},\mathbf{q}}|\theta_\gamma\rangle\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}H\right)^{\gamma}_{\ \alpha} + \frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\overrightarrow{\mathcal{D}}|\phi_\alpha\rangle\,\delta_{\mathbf{q},\mathbf{0}} - \left(S^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}|\phi_\gamma\rangle\left(S^{-1}H\right)^{\gamma}_{\ \alpha}\Bigg] \tag{4.4.45}$$

Total gradient

By combining the explicit and implicit forms of the first-order NGWF gradient, we obtain the rather compact expression:

$$|\left(g^{\delta\phi}_{\mathrm{EXP}}\right)^{\eta}\rangle = 2s\frac{\delta\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \alpha\beta}\langle\theta_\beta|}{\delta\langle\theta_\eta|}\Bigg[\widetilde{H}_{\mathbf{q},\mathbf{q}}|\theta_\gamma\rangle\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}S\right)^{\gamma}_{\ \alpha} - \hat{S}_{\mathbf{q},\mathbf{q}}|\theta_\gamma\rangle\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}H\right)^{\gamma}_{\ \alpha} + \frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\overrightarrow{\mathcal{D}}|\phi_\alpha\rangle\,\delta_{\mathbf{q},\mathbf{0}} + \left(\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}|\phi_\alpha\rangle - \left(S^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}|\phi_\gamma\rangle\left(S^{-1}H\right)^{\gamma}_{\ \alpha}\Bigg] \tag{4.4.46}$$

In order to obtain $\frac{\delta\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \alpha\beta}\langle\theta_\beta|}{\delta\langle\theta_\eta|}$, we turn to the gauge constraint relation (Eq. 4.4.16), along with the expression for the conduction kernel (Eq. 4.4.7) and the identity operator in the joint basis ($|\theta_\alpha\rangle\left(S^{\theta\ -1}\right)^{\alpha\beta}\langle\theta_\beta|\,\hat{S} = \hat{1}$):

$$\frac{\delta\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger\ \alpha\beta}\langle\theta_\beta|}{\delta\langle\theta_\eta|} = \left[KS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right]^{\alpha\eta} - \left[KS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right]^{\alpha\eta}\hat{S}_{\mathbf{q},\mathbf{q}}|\phi_\mu\rangle\,K^{\mu\nu}_{\mathbf{q}\mathbf{q}}\langle\phi_\nu| \tag{4.4.47}$$

where we have approximated the auxiliary response kernel $\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}$ to be independent of the joint basis θ. By defining a matrix $X_{\mathbf{q},\mathbf{0}}$ as:

$$X_{\mathbf{q},\mathbf{0}\ \alpha\beta} = \Big[\left(H^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger} + H^{\phi\theta}_{\mathbf{q},\mathbf{q}}P^{1\lambda_{\mathbf{q}}\ \theta\phi}S - S^{\phi\theta}_{\mathbf{q},\mathbf{q}}P^{1\lambda_{\mathbf{q}}\ \theta\phi}\,H + \frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\overrightarrow{\mathcal{D}}\delta_{\mathbf{q},\mathbf{0}} - \left(S^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}S^{-1}H\Big]_{\alpha\beta}, \tag{4.4.48}$$

we can plug Eqs. 4.4.47 and 4.4.48 into the total gradient expression (Eq. 4.4.46) and obtain the final expression for the total contra-variant gradient:

$$|\left(g^{\theta}\right)^{\eta}\rangle = 2s\ \widetilde{H}_{\mathbf{q},\mathbf{q}}\,|\theta_{\alpha}\rangle\left(P^{1\lambda_{\mathbf{q}}\ \theta\phi}S\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right)^{\alpha\eta} + 2s\ \widetilde{S}_{\mathbf{q},\mathbf{q}}\,|\theta_{\alpha}\rangle\left(-P^{1\lambda_{\mathbf{q}}\ \theta\phi}HKS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right)^{\alpha\eta} +$$
$$2s\ \hat{S}_{\mathbf{q},\mathbf{q}}\,|\phi_{\alpha}\rangle\left(-K_{\mathbf{q},\mathbf{q}}X_{\mathbf{q},\mathbf{0}}KS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right)^{\alpha\eta} + 2s\ \left(\hat{S}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}|\phi_{\alpha}\rangle\left(-S^{-1}HKS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right)^{\alpha\eta} +$$
$$2s\ \left(\widetilde{H}^{1\lambda_{-\mathbf{q}}}_{\mathbf{0},\mathbf{q}}\right)^{\dagger}|\phi_{\alpha}\rangle\left[KS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right]^{\alpha\eta} + 2s\ \frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\overrightarrow{\mathcal{D}}\,|\phi_{\alpha}\rangle\left[KS\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}\right)^{\dagger}\right]^{\alpha\eta}\delta_{\mathbf{q},\mathbf{0}} \qquad (4.4.49)$$

The conjugate gradient $\langle\left(g^{\theta}\right)^{\eta}|$ can be obtained in exactly the same manner as outline above, and it turns out that it is unsurprisingly equal to $|\left(g^{\theta}\right)^{\eta}\rangle^{\dagger}$. Note the use of both the auxiliary response kernel $\underline{P}^{1\lambda_{\mathbf{q}}\ \theta\phi}$ and the gauge-constrained response kernel $P^{1\lambda_{\mathbf{q}}\ \theta\phi}$ in Eq. 4.4.49.

In order to update the first-order NGWFs, one needs the covariant gradients $\langle\mathbf{r}|\left(g^{\theta}\right)_{\eta}\rangle$ and $\langle\mathbf{r}|\left(g^{\theta}\right)_{\eta}\rangle^{\dagger}$, which are obtained directly from the previously discussed contra-variant ones. Therefore, we aim to update the basis functions as:

$$\theta'(\mathbf{r}) = \theta(\mathbf{r}) + \left[g^{\theta}_{\eta}(\mathbf{r}) + (g^{\theta\ \dagger}_{\eta}(\mathbf{r}))^{\dagger}\right]\Delta = \theta(\mathbf{r}) + 2\ g^{\theta}_{\eta}(\mathbf{r})\Delta\ , \qquad (4.4.50)$$

where Δ is a line step. The individual covariant gradients (for each NGWF subset in the joint set) are obtained as $g^{\delta\phi}_{\eta}(\mathbf{r}) = \left(g^{\delta\phi}\right)^{\mu}(\mathbf{r})\left(S^{\delta_{\phi}}_{\mathbf{q},\mathbf{q}}\right)_{\mu\eta}$ and $g^{\phi}_{\eta}(\mathbf{r}) = \left(g^{\phi}\right)^{\mu}(\mathbf{r})\left(S_{\mathbf{q},\mathbf{q}}\right)_{\mu\eta}$.

However, the difficulty lies in the fact that we cannot update all the functions in the $\theta = \{\delta\phi, \phi_{\mathbf{q}}\}$ set, but only the $\{\delta\phi\}$ subset. Therefore, in order to get to the global solution faster, we need to use a covariant gradient that fully compensates for the change in the first-order waverunction, while only modifying $\{\delta\phi\}$. For this, one requires that:

$$|\phi + g^{\phi}_{\eta}\Delta\rangle\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \phi\phi}\right)^{\eta\alpha} + |\delta\phi + g^{\delta\phi}_{\eta}\Delta\rangle\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \delta\phi\phi}\right)^{\eta\alpha} \approx$$
$$|\phi_{\eta}\rangle\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \phi\phi}\right)^{\eta\alpha} + |\delta\phi_{\eta} + g^{\delta\phi'}_{\eta}\Delta\rangle\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \delta\phi\phi}\right)^{\eta\alpha} \qquad (4.4.51)$$

The exact solution to this predicament is to define the equivalent covariant gradient for the first-order NGWFs $\{\delta\phi\}$ as:

$$|g^{\delta\phi'}_{\eta}\rangle = |\left(g^{\delta\phi}\right)^{\mu}\rangle\left(S^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)_{\mu\eta} + |\left(g^{\phi}\right)^{\mu}\rangle\left(S_{\mathbf{q},\mathbf{q}}\right)_{\mu\nu}\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \phi\phi}\right)^{\nu\alpha}\left(\left[\underline{P}^{1\lambda_{\mathbf{q}}\ \delta\phi\phi}\right]^{-1}\right)_{\alpha\eta} \approx \qquad (4.4.52)$$
$$|\left(g^{\delta\phi}\right)^{\mu}\rangle\left(S^{\delta\phi}_{\mathbf{q},\mathbf{q}}\right)_{\mu\eta} + |\left(g^{\phi}\right)^{\mu}\rangle\left(S_{\mathbf{q},\mathbf{q}}\right)_{\mu\nu}\left(\underline{P}^{1\lambda_{\mathbf{q}}\ \phi\phi}\right)^{\nu\alpha}S_{\alpha\eta}\ ,$$

where we have used the approximation on the last line simply because we do not have access to the inverse of the $\underline{P}^{1\lambda_{\mathbf{q}}\ \delta\phi\phi}$ component of the auxiliary response kernel. Thus,

we can employ this approximation such that only $\{\delta\phi\}$ is modified, leaving the joint set $\{\theta\}$ as a common ground between first-order $\{\delta\phi\}$ and valence NGWFs $\{\phi\}$.

4.4.7 Perturbation basis transformation

Throughout the first-order wavefunction optimisation, if we perturb only one atom, we risk having zero-valued response NGWFs (Sec. 4.4.6) far away from the perturbation, which would lead to failures in the inversion of overlap matrices. Therefore, it is more efficient to perturb all the NGWFs at once, and initialise them with (the negative of) the rigid real-space gradient of the valence orbitals $\frac{\partial\phi_\alpha(\mathbf{r})}{\partial\mathbf{r}}$. In order to achieve this, one needs to transform the perturbation basis from a 1-atom 1-direction scheme (with weight 1) to an all-atoms 1-direction scheme, (with various weighting factors). For instance, the force-constant matrix would transform as

$$\Phi_{\mathbf{\Lambda}_i\mathbf{\Lambda}_j} = (O_{\Lambda\leftarrow\lambda})_{ik}\Phi_{\lambda_k\lambda_l}(O_{\lambda\leftarrow\Lambda})_{lj} \ , \tag{4.4.53}$$

where λ_i are perturbation-basis row-vectors in the 1-atom 1-direction scheme, $\mathbf{\Lambda}_i$ are vectors in the all-atoms 1-direction approach, and O is the transformation matrix of size $3\mathrm{N}_{\mathrm{atoms}} \times 3N_{\mathrm{atoms}}$. Analytically, we have the advantage that the 1-atom 1-direction perturbation-basis vectors are actually simply unit vectors, with only one non-zero quantity in their $3\mathrm{N}_{\mathrm{atoms}}$ entries. Thus, we have:

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ ... \\ \lambda_{3N_{\mathrm{atoms}}} \end{pmatrix} = I_{3N_{\mathrm{atoms}}} \ , \tag{4.4.54}$$

where $I_{3N_{\mathrm{atoms}}}$ is the identity matrix. Consequently, since we know that the inverse transformation matrix is defined as:

$$O_{\lambda\leftarrow\Lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ ... \\ \lambda_{3N_{\mathrm{atoms}}} \end{pmatrix} \left(\mathbf{\Lambda}_1^T|\mathbf{\Lambda}_2^T|...|\mathbf{\Lambda}_{3N_{\mathrm{atoms}}}^T\right) \ , \tag{4.4.55}$$

it is clear that the transformation matrix from the λ-basis to the Λ-basis can be defined as

$$O_{\Lambda\leftarrow\lambda} = \left(\mathbf{\Lambda}_1^T|\mathbf{\Lambda}_2^T|...|\mathbf{\Lambda}_{3N_{\mathrm{atoms}}}^T\right)^{-1} \tag{4.4.56}$$

Thus, it is practical to calculate the force-constant matrix in the compound Λ basis; if we require the force-constant matrix in the 1-atom 1-direction basis, we can simply transform back by using:

$$\Phi_{\lambda_i\lambda_j} = O_{ik}^{-1}\Phi_{\mathbf{\Lambda}_k\mathbf{\Lambda}_l}O_{lj} \ , \tag{4.4.57}$$

The only remaining difficulty is to construct this transformation matrix O. Our reasoning was that only the relative atom movements are relevant, since the studied systems will be periodic, so the vibrational frequencies are independent of cell translations. Thus, moving a certain atom by one unit in a direction is equivalent to moving the same atom by 0.5 units in the same direction, while moving all the others 0.5 units in the opposite direction. Therefore, as discussed in the beginning of the subsection, we manage to perturb all the ions at once, allowing for an efficient initialisation of the first-order NGWFs.

We have found that other seemingly equivalent approaches, such as products of rotation matrices, produce perturbation weights that vary significantly in magnitude between ions. While in principle a linear transformation should not affect the results at all, the NGWF CG optimisation induces a non-linear component. Therefore, the best transformation matrices are those that result in weights with similar absolute values for all moved atoms.

Lastly, we note that QuantumEspresso [105, 106, 107] also utilises collective atom movements, but it determines the weights by first performing a symmetry analysis to approximate the vibrational mode coefficients.

4.5 Polarisation for periodic systems

In this section we describe the theoretical formalism that has allowed us to calculate the dipole operator $\hat{D}$ and its associate quantities in the expression for the second-order energy (Eq .4.4.19).

4.5.1 Zeroth-order

As starting point we use the account of Kudin et al. [110] on the modern theory of polarisation in systems with local orbitals (such as our NGWFs), to which we have also

added PAW correction terms. Thus, the polarisation vector is defined as:

$$\mathbf{P} = \mathbf{P}_I - s\frac{V}{(2\pi)^3}(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2) \ , \tag{4.5.1}$$

where s is the spin factor, V is the cell volume, and $\mathbf{P}_I$ is the ion-induced polarisation:

$$\mathbf{P}_I = \sum_i Z_i(\mathbf{R}_i - \mathbf{R}_c) \ , \tag{4.5.2}$$

with $\mathbf{R}_c = \sum_j \mathbf{R}_j Z_j / \sum_l Z_l$ being a reference position which we call the centre of ionic charge.

Starting with $\boldsymbol{\gamma}_2$, it is described by:

$$\boldsymbol{\gamma}_2 = i\sum_n^{N/2} \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k} M_n^{\dagger\ \mu}(\mathbf{k}) S_{\mu\nu}(\mathbf{k}) \frac{d}{d\mathbf{k}} M^{\nu}{}_n(\mathbf{k}) = -\mathrm{Im}\Big\{\ln \prod_{j=1}^{N_k} \det\Big[M^{\dagger}(\mathbf{k}_j) S(\mathbf{k}_j) M(\mathbf{k}_{j+1})\Big]\Big\} \tag{4.5.3}$$

In the previous expression, $M^{\nu}{}_n(\mathbf{k})$ are molecular-orbital elements obtained from the definition of the pseudo-eigenvector ($\mathbf{R}$ are multiples of the real-space lattice vectors):

$$\widetilde{\psi}_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_\nu(\mathbf{r} - \mathbf{R}) M^{\nu}{}_n(\mathbf{k}) \tag{4.5.4}$$

If one recalls that $\hat{S}$ is the PAW overlap operator (Eq. 2.2.6), O_{ab} is the partial-wave overlap (Eq. 2.2.6), and if we define $\phi_\nu^{\mathbf{R}}(\mathbf{r})$ to be the duplicate of $\phi_\nu(\mathbf{r})$ in a periodic cell displaced by $\mathbf{R}$ from the origin, then the overlap matrix between the NGWFs is:

$$\begin{aligned} S_{\mu\nu}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \phi_\mu | \hat{S} | \phi_\nu^{\mathbf{R}} \rangle = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \Big[\int \phi_\mu^*(\mathbf{r}) \phi_\nu(\mathbf{r} - \mathbf{R}) \ \mathrm{d}\mathbf{r} + \\ \sum_{a,b} \Big(\int \phi_\mu^*(\mathbf{r}) \widetilde{p}^a(\mathbf{r}) \ \mathrm{d}\mathbf{r} \Big) O_{ab} \Big(\int \ \widetilde{p}^b(\mathbf{r}') \phi_\nu(\mathbf{r}' - \mathbf{R}) \mathrm{d}\mathbf{r}' \Big) \Big] \end{aligned} \tag{4.5.5}$$

Also, N_k is the number of discretised k-points in the Brillouin Zone (BZ), meaning that $\mathbf{k}_{N_k+1} = \mathbf{k}_1$. This offers us relief in the case of Γ-point sampling only (where $N_k = 1$ and $\mathbf{k} = \mathbf{0}$), because it implies that:

$$\boldsymbol{\gamma}_2 = -\mathrm{Im}\Big\{\ln \det\Big[M^{\dagger}(\mathbf{0}) S(\mathbf{0}) M(\mathbf{0})\Big]\Big\} = -\mathrm{Im}\Big\{\ln \det\Big[\mathrm{Id}\Big]\Big\} = 0 \ , \tag{4.5.6}$$

where we have used the orthonormality of the wavefunctions belonging to the same k-point. Thus, for Γ-point calculations, the $\boldsymbol{\gamma}_2$ contribution is of no concern.

As for $\boldsymbol{\gamma}_1$, one has:

$$\boldsymbol{\gamma}_1 = \sum_n^{N/2} \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k} M_n^{\dagger\ \mu}(\mathbf{k}) \Big[\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \Big(\mathbf{Z}_{\mu\nu}^{\mathbf{0R}} - \mathbf{R} S_{\mu\nu}^{\mathbf{0R}} \Big) \Big] M^{\nu}{}_n(\mathbf{k}) \ , \tag{4.5.7}$$

where $\mathbf{R}$ is once again a multiple of lattice vectors. We also require the definitions

$$\begin{aligned} S_{\mu\nu}^{\mathbf{0R}} = \langle \phi_\mu | \hat{S} | \phi_\nu^{\mathbf{R}} \rangle = \int \phi_\mu(\mathbf{r}) \phi_\nu(\mathbf{r} - \mathbf{R}) \ \mathrm{d}\mathbf{r} + \\ \sum_{a,b} \Big(\int \phi_\mu(\mathbf{r}) \widetilde{p}^a(\mathbf{r}) \ \mathrm{d}\mathbf{r} \Big) O_{ab} \Big(\int \ \widetilde{p}^b(\mathbf{r}') \phi_\nu(\mathbf{r}' - \mathbf{R}) \mathrm{d}\mathbf{r}' \Big) \end{aligned} \tag{4.5.8}$$

and

$$\begin{aligned} \mathbf{Z}_{\mu\nu}^{\mathbf{0R}} = \langle \phi_\mu | \hat{\tau}^\dagger(\mathbf{r} - \mathbf{R}_c) \hat{\tau} | \phi_\nu^{\mathbf{R}} \rangle = \int \phi_\mu(\mathbf{r}) (\mathbf{r} - \mathbf{R}_c) \phi_\nu(\mathbf{r} - \mathbf{R}) \ \mathrm{d}\mathbf{r} + \sum_{a,b} \Bigg\{ \int \phi_\mu(\mathbf{r}) \widetilde{p}^a(\mathbf{r}) \ \mathrm{d}\mathbf{r} \\ \Big[\langle \varphi_a | (\mathbf{r} - \mathbf{R}_c) | \varphi_b \rangle - \langle \widetilde{\varphi}_a | (\mathbf{r} - \mathbf{R}_c) | \widetilde{\varphi}_b \rangle \Big] \int \ \widetilde{p}^b(\mathbf{r}') \phi_\nu(\mathbf{r}' - \mathbf{R}) \mathrm{d}\mathbf{r}' \Bigg\} = \\ - \mathbf{R}_c S_{\mu\nu}^{\mathbf{0R}} + \int \phi_\mu(\mathbf{r}) \ \mathbf{r} \ \phi_\nu(\mathbf{r} - \mathbf{R}) \ \mathrm{d}\mathbf{r} + \sum_{a,b} \Bigg\{ \int \phi_\mu(\mathbf{r}) \widetilde{p}^a(\mathbf{r}) \ \mathrm{d}\mathbf{r} \\ \Big[\langle \varphi_a | \mathbf{r} | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \mathbf{r} | \widetilde{\varphi}_b \rangle \Big) \int \ \widetilde{p}^b(\mathbf{r}') \phi_\nu(\mathbf{r}' - \mathbf{R}) \mathrm{d}\mathbf{r}' \Bigg\} \end{aligned} \tag{4.5.9}$$

where $\mathbf{r}$ is the position vector with respect to the cell origin. However, for practical implementation purposes we wish to have the NGWFs from outside the original cell being as the "bra" term. Thus, by switching the NGWFs around in Eq. 4.5.8, we clearly see that $S_{\mu\nu}^{\mathbf{0R}} = S_{\nu\mu}^{\mathbf{R0}}$ at the Γ point (where orbitals are real):

$$\begin{aligned} S_{\nu\mu}^{\mathbf{R0}} = \langle \phi_\nu^{\mathbf{R}} | \hat{S} | \phi_\mu \rangle = \int \phi_\nu(\mathbf{r} - \mathbf{R}) \phi_\mu(\mathbf{r}) \ \mathrm{d}\mathbf{r} + \\ \sum_{a,b} \Big(\int \phi_\nu(\mathbf{r} - \mathbf{R}) \widetilde{p}^a(\mathbf{r}) \ \mathrm{d}\mathbf{r} \Big) O_{ab} \Big(\int \ \widetilde{p}^b(\mathbf{r}') \phi_\mu(\mathbf{r}') \mathrm{d}\mathbf{r}' \Big) \end{aligned} \tag{4.5.10}$$

Similarly, $\mathbf{Z}_{\mu\nu}^{\mathbf{0R}} = \mathbf{Z}_{\nu\mu}^{\mathbf{R0}}$. Furthermore, the ONETEP implementation of overlaps between NGWFs is based on the dissociation of $\mathbf{r}$ into $\mathbf{r}_\mu + \mathbf{R}_\mu^{\mathrm{FFT}}$, where $\mathbf{r}_\mu$ is with respect to center of function (FFT box) μ, while $\mathbf{R}_\mu^{\mathrm{FFT}}$ is the position of the function (FFT box) with respect to the cell origin. Similarly, in the integral between partial waves, $\mathbf{r}$ is split into $\mathbf{R}_{I(b)} + \mathbf{r}_b$, where $\mathbf{R}_{I(b)}$ is the position of the associated atom, while $\mathbf{r}_b$ is with respect to the center of that atom. Consequently, one has

$$\mathbf{Z}_{\nu\mu}^{\mathbf{R0}} = - \mathbf{R}_c S_{\nu\mu}^{\mathbf{R0}} + \mathbf{R}_\mu^{\mathrm{FFT}} \int \phi_\nu(\mathbf{r} - \mathbf{R}) \phi_\mu(\mathbf{r}) \ \mathrm{d}\mathbf{r} + \int \phi_\nu(\mathbf{r} - \mathbf{R}) \ \mathbf{r}_\mu \ \phi_\mu(\mathbf{r}) \ \mathrm{d}\mathbf{r} + \tag{4.5.11}$$

$$\sum_{a,b} \mathbf{R}_{I(b)} \Big(\int \phi_\nu(\mathbf{r}-\mathbf{R}) \widetilde{p}^a(\mathbf{r}) \, \mathrm{d}\mathbf{r} \Big) O_{ab} \Big(\int \, \widetilde{p}^b(\mathbf{r}') \phi_\mu(\mathbf{r}') \mathrm{d}\mathbf{r}' \Big) +$$
$$\sum_{a,b} \int \phi_\nu(\mathbf{r}-\mathbf{R}) \widetilde{p}^a(\mathbf{r}) \, \mathrm{d}\mathbf{r} \Big[\, \langle \varphi_a | \mathbf{r}_b | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \mathbf{r}_b | \widetilde{\varphi}_b \rangle \, \Big] \int \, \widetilde{p}^b(\mathbf{r}') \phi_\mu(\mathbf{r}') \mathrm{d}\mathbf{r}'$$

However, it is clear that $\mathbf{Z}^{\mathbf{R0}}_{\nu\mu}$ terms do not have any explicit phase associated to the cell in which the calculation is done, so we can use the existing ONETEP subroutines and simply calculate:

$$\mathbf{Z}_{\nu\mu} = \sum_{\mathbf{R}} \mathbf{Z}^{\mathbf{R0}}_{\nu\mu} = -\mathbf{R}_c S_{\nu\mu} + \mathbf{R}^{\mathrm{FFT}}_{\mu} \langle \phi_\nu | \phi_\mu \rangle + \langle \phi_\nu | \mathbf{r}_\mu | \phi_\mu \rangle + \sum_I \mathbf{R}_I \sum_{a,b \in I} \langle \phi_\nu | \widetilde{p}^a \rangle \, O_{ab} \, \langle \widetilde{p}^b | \phi_\mu \rangle +$$
$$\sum_I \sum_{a,b \in I} \langle \phi_\nu | \widetilde{p}^a \rangle \Big(\, \langle \varphi_a | \mathbf{r}_b | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \mathbf{r}_b | \widetilde{\varphi}_b \rangle \, \Big) \, \langle \widetilde{p}^b | \phi_\mu \rangle \tag{4.5.12}$$

Considering that we only utilise the Γ-point ground state wavefunctions, meaning that the eigenvectors (M_n) are real and that $K^{\mu\nu} = K^{\nu\mu}$ is real and symmetric, one finally has:

$$\mathbf{P} = \sum_I Z_I (\mathbf{R}_I - \mathbf{R}_c) - s \Big[\mathbf{Z}_{\nu\mu} - \sum_{\mathbf{R}} \mathbf{R} S^{\mathbf{R0}}_{\nu\mu} \Big] K^{\mu\nu} \tag{4.5.13}$$

Fortunately, due to the fact that the NGWFs cannot exceed the simulation cell size, we can limit the sum over $\mathbf{R}$ to only the nearest neighbours (i.e. 27 terms, out of which most will be 0).

After some manipulation, one can write the polarisation by defining a dipole operator vector:

$$\hat{\mathcal{D}} = \hat{\tau}^\dagger \Big[\int |\mathbf{r}\rangle \, \mathbf{r} \, \langle \mathbf{r}| \, \, \mathrm{d}\mathbf{r} \Big] \hat{\tau} - \mathbf{R}_c \hat{S} - \sum_{\mathbf{R}} \Big[\int |\mathbf{r} - \mathbf{R}\rangle \, \mathbf{R} \, \langle \mathbf{r}| \, \, \mathrm{d}\mathbf{r} \Big] \hat{S} \, , \tag{4.5.14}$$

where $\hat{\tau}$ is the PAW projection operator (Eq. 2.2.1), $\hat{S}$ is the PAW overlap operator (Eq. 2.2.6), $\mathbf{R}_c$ is the centre of ionic charge, and $\mathbf{R}$ is a multiple of the lattice vectors. Therefore, we can write the polarisation in an elegant manner:

$$\mathbf{P} = \sum_I Z_I (\mathbf{R}_I - \mathbf{R}_c) - s \, \langle \phi_\nu | \, \hat{\mathcal{D}} \, | \phi_\mu \rangle \, K^{\mu\nu} = \sum_I Z_I (\mathbf{R}_I - \mathbf{R}_c) - s \sum_{i=1}^{N_e/2} \langle \psi_i | \, \hat{\mathcal{D}} \, | \psi_i \rangle \tag{4.5.15}$$

4.5.2 First-order polarisation: Born charges

The Born charges ($\mathbf{Z}^{\mathrm{B}}_{\boldsymbol{\lambda}_\mathbf{0}}$) are defined as the change in polarisation due to the perturbation of atoms at $\mathbf{q} = \mathbf{0}$. Starting with the nuclear term, and remembering that the center of

ionic charge is $\mathbf{R}_c = \sum_j \mathbf{R}_j Z_j / \sum_l Z_l$, we can show that it is conveniently null:

$$\frac{d}{d\lambda_{\mathbf{0}}} \sum_i Z_i(\mathbf{R}_i - \mathbf{R}_c) = \sum_{\mathbf{R}} Z_i \mathbf{u}_{\lambda(i)} w_{\lambda(i)} - \sum_{\mathbf{R}} \Big(\sum_i Z_i \Big) \frac{\sum_j Z_j \mathbf{u}_{\lambda(j)} w_{\lambda(j)}}{\sum_l Z_l} = 0 \quad (4.5.16)$$

where $\mathbf{u}_{\lambda(i)}$ is the direction of perturbation λ that is imposed on atom i, while $w_{\lambda(i)}$ is the weight of the aforementioned quantity. Therefore, we are left only with the electronic term, which in the limiting case of Γ-point ground-state calculations is:

$$\begin{aligned} \mathbf{Z}^{\mathrm{B}}_{\lambda_{\mathbf{0}}} = -s \frac{d}{d\lambda_{\mathbf{0}}} \Big[\mathbf{Z}_{\nu\mu} - \sum_{\mathbf{R}} \mathbf{R} S^{\mathbf{R0}}_{\nu\mu} \Big] K^{\mu\nu} = -s \Big[\mathbf{Z}^{\theta\phi}_{\nu\mu -} \sum_{\mathbf{R}} \mathbf{R} S^{\mathbf{R0}\ \theta\phi}_{\nu\mu} \Big] \Big(P^{1\lambda_{\mathbf{0}}\ \theta\phi\ \dagger} \Big)^{\mu\nu} - \\ s \Big[\mathbf{Z}^{\phi\theta}_{\nu\mu} - \sum_{\mathbf{R}} \mathbf{R} S^{\mathbf{R0}\ \phi\theta}_{\nu\mu} \Big] \Big(P^{1\lambda_{\mathbf{0}}\ \theta\phi} \Big)^{\mu\nu} - s \Big[\mathbf{Z}^{1\lambda_{\mathbf{0}}}_{\nu\mu} - \sum_{\mathbf{R}} \mathbf{R} \big(S^{1\lambda_{\mathbf{0}}\ \mathbf{R0}}_{\nu\mu} \big) \Big] K^{\mu\nu} \ , \end{aligned} \quad (4.5.17)$$

where the first order terms contain perturbations of the ionic charge center and projectors:

$$\begin{aligned} S^{1\lambda_{\mathbf{0}}\ \mathbf{R0}}_{\nu\mu} = & \sum_I \sum_{a,b \in I} \Big(\int \phi_\nu(\mathbf{r} - \mathbf{R}) \frac{\partial \widetilde{p}^a(\mathbf{r})}{\partial \lambda_{\mathbf{0}}(I)} \, \mathrm{d}\mathbf{r} \Big) O_{ab} \Big(\int \widetilde{p}^b(\mathbf{r}') \phi_\mu(\mathbf{r}') \mathrm{d}\mathbf{r}' \Big) + \\ & \sum_I \sum_{a,b \in I} \Big(\int \phi_\nu(\mathbf{r} - \mathbf{R}) \widetilde{p}^a(\mathbf{r}) \, \mathrm{d}\mathbf{r} \Big) O_{ab} \Big(\int \frac{\partial \widetilde{p}^b(\mathbf{r}')}{\partial \lambda_{\mathbf{0}}(I)} \phi_\mu(\mathbf{r}') \mathrm{d}\mathbf{r}' \Big) \end{aligned} \quad (4.5.18)$$

and

$$\begin{aligned} \mathbf{Z}^{1\lambda_{\mathbf{0}}}_{\nu\mu} = & - \sum_I \frac{Z_I}{\sum_J Z_J} \mathbf{u}_{\lambda_{\mathbf{0}}(I)} w_{\lambda_{\mathbf{0}}(I)} S_{\nu\mu} - \mathbf{R}_c S^{1\lambda_{\mathbf{0}}}_{\nu\mu} + \sum_I \mathbf{R}_I \sum_{a,b \in I} \Bigg[\langle \phi_\nu | \frac{\partial \widetilde{p}^a}{\partial \lambda_{\mathbf{0}}(I)} \rangle \, O_{ab} \, \langle \widetilde{p}^b | \phi_\mu \rangle + \\ & \langle \phi_\nu | \widetilde{p}^a \rangle \, O_{ab} \, \langle \frac{\partial \widetilde{p}^b}{\partial \lambda_{\mathbf{0}}(I)} | \phi_\mu \rangle \Bigg] + \sum_I \sum_{a,b \in I} \Bigg[\langle \phi_\nu | \frac{\partial \widetilde{p}^a}{\partial \lambda_{\mathbf{0}}(I)} \rangle \Big(\langle \varphi_a | \mathbf{r}_b | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \mathbf{r}_b | \widetilde{\varphi}_b \rangle \Big) \langle \widetilde{p}^b | \phi_\mu \rangle + \\ & \langle \phi_\nu | \widetilde{p}^a \rangle \Big(\langle \varphi_a | \mathbf{r}_b | \varphi_b \rangle - \langle \widetilde{\varphi}_a | \mathbf{r}_b | \widetilde{\varphi}_b \rangle \Big) \langle \frac{\partial \widetilde{p}^b}{\partial \lambda_{\mathbf{0}}(I)} | \phi_\mu \rangle \Bigg] + \sum_I \mathbf{u}_{\lambda_{\mathbf{0}}(I)} w_{\lambda_{\mathbf{0}}(I)} \sum_{a,b \in I} \langle \phi_\nu | \widetilde{p}^a \rangle \, O_{ab} \, \langle \widetilde{p}^b | \phi_\mu \rangle \end{aligned} \quad (4.5.19)$$

4.5.3 First-order polarisation: electric field perturbation

In the case of electric field perturbations, the formula for the first-order polarisation is the one described in Eq. 4.5.17, with the only difference that $\mathbf{Z}^{1\epsilon_{\mathbf{0}}}_{\nu\mu} = \mathbf{0}$ and $S^{1\epsilon_{\mathbf{0}}\ \mathbf{R0}}_{\nu\mu} = 0$ (i.e. $\hat{\mathcal{D}}^{1\epsilon_{\mathbf{0}}} = \mathbf{0}$), as no projectors or ions are moved:

$$\frac{d\mathbf{P}}{d\epsilon_{\mathbf{0}}} = - \, s \, \langle \theta_\nu | \, \hat{\mathcal{D}} \, | \phi_\mu \rangle \Big(P^{1\epsilon_{\mathbf{0}}\ \theta\phi} \Big)^{\dagger\ \mu\nu} - s \, \langle \phi_\nu | \, \hat{\mathcal{D}} \, | \theta_\mu \rangle \Big(P^{1\epsilon_{\mathbf{0}}\ \theta\phi} \Big)^{\mu\nu} = \quad (4.5.20)$$

$$-\sum_{i=1}^{N_e/2} f_i \langle \psi_i^{1\epsilon_0}| \hat{\mathcal{D}} |\psi_i\rangle - \sum_{i=1}^{N_e/2} f_i \langle \psi_i| \hat{\mathcal{D}} |\psi_i^{1\epsilon_0}\rangle$$

Since in the derivation of the Sternheimer Equation for electric field perturbations (Eq. 4.2.38) we need the $\frac{\partial}{\partial \langle \widetilde{\psi}_i^{1\epsilon_0}|}$ of the previously-derived first-order polarisation, it is clear from Eq. 4.5.20 that:

$$\frac{\partial}{\partial \langle \widetilde{\psi}_i^{1\epsilon_0}|} \frac{d\mathbf{P}}{d\epsilon_{\mathbf{0}}} = -f_i \hat{\mathcal{D}} |\widetilde{\psi}_i\rangle \ , \tag{4.5.21}$$

where $\hat{\mathcal{D}}$ was defined in Eq. 4.5.14. Notice that we have only considered $\mathbf{q} = \mathbf{0}$ since this is the only q-vector needed to obtain the dielectric matrix.

4.6 Solving for the response kernel

In this section we elaborate on how we solve the equation governing the auxiliary response kernel ($\mathbf{X}^{\theta\phi}$ in Eq. 4.4.23). As previously explained, this takes the form of a generalised Sylvester equation, with the added complication that $\mathbf{E}^{\theta\phi}$ also depends on the auxiliary response $\mathbf{X}^{\theta\phi}$.

Our algorithm combines elements from a number of works by Chehab et al. [111, 112, 113] and other authors [114, 115, 116, 117, 118, 119]. However, we warn the readers that in Ref. [113] there are several mistakes, which have been corrected in this thesis. We start by defining a linear matrix operator:

$$\mathcal{F}(\mathbf{X}^{\theta\phi}) = \mathcal{A}(\mathbf{X}^{\theta\phi}) - \mathbf{E}^{\theta\phi} \ , \tag{4.6.1}$$

with

$$\mathcal{A}(\mathbf{X}^{\theta\phi}) = \mathbf{A}^{\theta}\mathbf{X}^{\theta\phi}\mathbf{B} + \mathbf{C}^{\theta}\mathbf{X}^{\theta\phi}\mathbf{D} \ , \tag{4.6.2}$$

where the involved matrices have been defined in Eq. 4.4.23 and 4.4.24. By optimising the response kernel we aim to minimise the residual of the equation $\mathcal{F}(\mathbf{X}^{\theta\phi}) = 0$, or equivalently, the square of the Frobenius norm of $\mathcal{F}(\mathbf{X}^{\theta\phi})$:

$$f(\mathbf{X}^{\theta\phi}) = ||\mathcal{F}(\mathbf{X}^{\theta\phi})||_F^2 = \langle \ \mathcal{F}(\mathbf{X}^{\theta\phi}), \mathcal{F}(\mathbf{X}^{\theta\phi}) \ \rangle = \mathrm{Tr}\Big\{ \big[\mathcal{F}(\mathbf{X}^{\theta\phi})\big]^{\dagger} \mathcal{F}(\mathbf{X}^{\theta\phi}) \Big\} \tag{4.6.3}$$

The adjoint of a linear operator ($\mathcal{A}^{\dagger}$) is generally defined such that $\langle \mathcal{A}^{\dagger}(\mathbf{X}), \mathbf{X} \rangle = \langle \mathbf{X}, \mathcal{A}(\mathbf{X}) \rangle$. By applying that in the context of the article of Chehab et al. [113], we can extend his

formula for the gradient of $f(\mathbf{X}^{\theta\phi})$ to the case of complex matrices, thus obtaining:

$$\nabla f(\mathbf{X}^{\theta\phi}) = 2\mathrm{Re}\Big[\Big(\mathbf{A}^{\theta}\Big)^{\dagger}\mathcal{F}(\mathbf{X}^{\theta\phi})\ \mathbf{B}^{\dagger} + \Big(\mathbf{C}^{\theta}\Big)^{\dagger}\mathcal{F}(\mathbf{X}^{\theta\phi})\ \mathbf{D}^{\dagger}\Big]. \tag{4.6.4}$$

This gradient will be useful in the line-search formula of the algorithm.

Since we aim for the linear-response phonon formalism to be as fast as possible, we are required to find a suitable preconditioning strategy for our algorithm. Thus, by using the Newton method, the iterations take the form:

$$\mathbf{X}^{\theta\phi}_{k+1} = \mathbf{X}^{\theta\phi}_{k} - \alpha_k\lambda_k \underbrace{\mathcal{F}'(\mathbf{X}^{\theta\phi}_{k})^{-1}\mathcal{F}(\mathbf{X}^{\theta\phi}_{k})}_{\mathbf{Z}^{\theta\phi}_{k}}\ , \tag{4.6.5}$$

where $\mathcal{F}'(\mathbf{X}^{\theta\phi}_{k})$ is the Fréchet (i.e. a generalised) derivative of $\mathcal{F}$ at $\mathbf{X}^{\theta\phi}_{k}$, $\mathbf{Z}^{\theta\phi}_{k}$ is the preconditioned search direction, α_k is a back-tracking constant (as explained later) and λ_k is the line step, as proposed by Glunt et al. [116] and Luengo et al. [119]:

$$\lambda_k = \lambda_{k-1}\left|\frac{\langle \mathbf{Z}^{\theta\phi}_{k-1}, \mathcal{F}(\mathbf{X}^{\theta\phi}_{k-1})\rangle}{\langle \mathbf{Z}^{\theta\phi}_{k-1}, \mathcal{F}(\mathbf{X}^{\theta\phi}_{k}) - \mathcal{F}(\mathbf{X}^{\theta\phi}_{k-1})\rangle}\right| \tag{4.6.6}$$

Since analytically inverting the Fréchet derivative is not an option, we must build a suitable approximation to the search direction $\mathbf{Z}^{\theta\phi}_{k}$.

4.6.1 Preconditioned and constrained search direction

Using the approach of Chehab et al. [112, 113], the preconditioned direction is the steady state of:

$$\frac{d\mathbf{Z}^{\theta\phi}_{k}}{dt} = \mathcal{F}(\mathbf{X}^{\theta\phi}_{k}) - \mathcal{F}'(\mathbf{X}^{\theta\phi}_{k})\mathbf{Z}^{\theta\phi}_{k}\ , \tag{4.6.7}$$

obtainable through a time marching scheme. Thus, we approximate $\mathcal{F}'(\mathbf{X}^{\theta\phi}_{k})\mathbf{Z}^{\theta\phi}_{k}$ through a finite-differencing scheme, but the result turns out to be independent of any infinitesimal differencing parameter τ:

$$\mathcal{F}'(\mathbf{X}^{\theta\phi}_{k})\mathbf{Z}^{\theta\phi}_{k} = \frac{\mathcal{F}(\mathbf{X}^{\theta\phi}_{k} + \tau\mathbf{Z}^{\theta\phi}_{k}) - \mathcal{F}(\mathbf{X}^{\theta\phi}_{k})}{\tau} = \mathcal{A}(\mathbf{Z}^{\theta\phi}_{k}) \tag{4.6.8}$$

Therefore, the ordinary differential equation (ODE) that needs to be solved is:

$$\frac{d\mathbf{Z}^{\theta\phi}_{k}}{dt} = \mathcal{F}(\mathbf{X}^{\theta\phi}_{k}) - \mathcal{A}(\mathbf{Z}^{\theta\phi}_{k}) = \mathcal{F}(\mathbf{X}^{\theta\phi}_{k} - \mathbf{Z}^{\theta\phi}_{k})\ , \tag{4.6.9}$$

for which we employ the "Enhanced-Cauchy 2" time marching scheme [113]:

1. Given the old search direction $\mathbf{Z}^{\theta\phi}_{k-1}$, copy it to a temporary matrix $\mathbf{Y}^{\theta\phi}_{1}$

2. For $p = 1, N_{\text{prec}}$ (N_{prec} is the maximum number of preconditioning iterations), define $\mathbf{G}^{\theta\phi}_{k,p} = \mathcal{F}\Big(\mathbf{X}^{\theta\phi}_{k} - \mathbf{Y}^{\theta\phi}_{p}\Big)$ and do:

   - Calculate $\mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big)$, $\mathcal{A}^2\big(\mathbf{G}^{\theta\phi}_{k,p}\big) = \mathcal{A}\Big(\mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\Big)$ and $\mathcal{A}^3\Big(\mathbf{G}^{\theta\phi}_{k,p}\Big)$

   - Calculate the following set of real parameters:

$$
\begin{aligned}
a &= \mathrm{Re}\,\langle \mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\rangle & b &= \mathrm{Re}\,\langle \mathcal{A}^2\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathcal{A}^2\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\rangle \\
c &= \mathrm{Re}\,\langle \mathcal{A}^3\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathcal{A}^3\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\rangle & d &= \mathrm{Re}\,\langle \mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathbf{G}^{\theta\phi}_{k,p}\rangle \\
e &= \mathrm{Re}\,\langle \mathcal{A}^2\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathbf{G}^{\theta\phi}_{k,p}\rangle & f &= \mathrm{Re}\,\langle \mathcal{A}^3\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathbf{G}^{\theta\phi}_{k,p}\rangle \\
g &= \mathrm{Re}\,\langle \mathcal{A}^2\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\rangle & \mu &= \mathrm{Re}\,\langle \mathcal{A}^3\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathcal{A}\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\rangle \\
& & \nu &= \mathrm{Re}\,\langle \mathcal{A}^3\big(\mathbf{G}^{\theta\phi}_{k,p}\big), \mathcal{A}^2\big(\mathbf{G}^{\theta\phi}_{k,p}\big)\rangle
\end{aligned}
$$

   - Advance one time step:

$$
\begin{aligned}
\mathbf{M}^{\theta\phi}_{1} &= \mathbf{G}^{\theta\phi}_{k,p} = \mathcal{F}(\mathbf{X}^{\theta\phi}_{k} - \mathbf{Y}^{\theta\phi}_{p}) \\
\mathbf{M}^{\theta\phi}_{2} &= \mathcal{F}(\mathbf{X}^{\theta\phi}_{k} - \mathbf{Y}^{\theta\phi}_{p} - \Delta t\mathbf{M}^{\theta\phi}_{1}) \\
\mathbf{M}^{\theta\phi}_{3} &= \mathcal{F}(\mathbf{X}^{\theta\phi}_{k} - \mathbf{Y}^{\theta\phi}_{p} - \Delta t\mathbf{M}^{\theta\phi}_{2}) \\
\mathbf{Y}^{\theta\phi}_{p+1} &= \mathbf{Y}^{\theta\phi}_{p} + \sum_{i=1}^{3} \beta_i \Delta t \mathbf{M}^{\theta\phi}_{i}
\end{aligned}
$$

   where

$$
\begin{aligned}
\Delta t &= \frac{-\mu\nu e - g\nu f + \mu f b + d\nu^2 - dcb + gce}{g^2 c + \mu^2 b - acb + a\nu^2 - 2\mu\nu g} \\
\beta_1 &= \frac{\nu f - \nu\Delta t\mu + \Delta t^2\nu^2 - \Delta t^2 bc - ec + \Delta t gc}{-\Delta t^2 bc + \Delta t^2\nu^2} \\
\beta_2 &= -\frac{\Delta t\nu f - \mu\Delta t^2\nu - \Delta t ec + \Delta t^2 gc - fb + \Delta t\mu b + \nu e - \nu\Delta t g}{-\Delta t^3 bc + \Delta t^3\nu^2} \\
\beta_3 &= 1 - \beta_1 - \beta_2
\end{aligned}
$$

   These expressions for these quantities ($\Delta t, \beta_1, \beta_2, \beta_3$) can be demonstrated by minimising the norm of the updated residual with respect to their values.

3. When the iterations are finished, $\mathbf{Y}^{\theta\phi}_{N_{\text{prec}}}$ becomes the new preconditioned direction $\mathbf{Z}^{\theta\phi}_{k+1}$

4.6.2 The full algorithm

We can now present the full algorithm for solving the generalised Sylvester equation of Eq. 4.4.23, and we will comment afterwards on some of the steps involved and the parameters used.

For $k = 1, N$ (N is the maximum number of iterations), do:

1. Update $\mathbf{E}^{\theta\phi}$, as a part of it depends on the response kernel. For this one first needs to construct the response kernel $\mathbf{P}^{1\lambda\ \theta\phi}$ from the auxiliary kernel $\mathbf{X}^{\theta\phi}$ by using 4.4.16, and then utilise $\mathbf{P}^{1\lambda\ \theta\phi}$ to rebuild the first-order densities and first-order hamiltonian $\mathbf{H}^{1\lambda-\mathbf{q}}_{\mathbf{0},\mathbf{q}}$. Construct $\mathcal{F}(\mathbf{X}^{\theta\phi}_k)$, $f(\mathbf{X}^{\theta\phi}_k)$ and $\nabla f(\mathbf{X}^{\theta\phi}_k)$, according to Eqs. 4.6.1, 4.6.3, and 4.6.4. If $\sqrt{f(\mathbf{X}^{\theta\phi}_k)}$ is sufficiently small the calculation is converged and the algorithm is stopped.

2. If $k = 1$, the line step is $\lambda_k = \lambda_0$.

3. If $k > 1$, use Eq. 4.6.6 to determine the line-step λ_k. Check if $\lambda_k \in [\sigma_{\min}; \sigma_{\max}]$; if not, set λ_k to be equal to the closest of $\sigma_{\min}$ and $\sigma_{\max}$.

4. Build the preconditioned direction $\mathbf{Z}^{\theta\phi}_k$, using the steps from subsection 4.6.1.

5. Set the initial value of the backtracking parameter $\alpha = \alpha_0$

6. Start the backtracking procedure:

   - If $|f(\mathbf{X}^{\theta\phi}_k - \lambda_k\alpha\mathbf{Z}^{\theta\phi}_k)| < |f(\mathbf{X}^{\theta\phi}_k) - \gamma\alpha\lambda_k \langle\nabla f(\mathbf{X}^{\theta\phi}_k), \mathbf{Z}\rangle|$ then set $\alpha_k = \alpha$, $\mathbf{X}^{\theta\phi}_{k+1} = \mathbf{X}^{\theta\phi}_k - \lambda_k\alpha_k\mathbf{Z}^{\theta\phi}_k$, $\lambda_k = \lambda_k\alpha_k$ and move onto $k = k + 1$.

   - Else, quadratically interpolate the backtracking parameter:

$$\alpha_{\text{new}} = \frac{\lambda_k}{2}\,\frac{\alpha^2|\langle\nabla f(\mathbf{X}^{\theta\phi}_k), \mathbf{Z}^{\theta\phi}_k\rangle|}{|f(\mathbf{X}^{\theta\phi}_k - \alpha\lambda_k\mathbf{Z}^{\theta\phi}_k) - f(\mathbf{X}^{\theta\phi}_k) + \lambda_k\alpha\langle\nabla f(\mathbf{X}^{\theta\phi}_k), \mathbf{Z}^{\theta\phi}_k\rangle|}$$

     – If $(\alpha_{\text{new}} < \tau_1\alpha)$ and $(\alpha_{\text{new}} > \tau_2\alpha)$, set $\alpha = \alpha_{\text{new}}$ and repeat step 6

     – Else, if the interpolation fails, set $\alpha = \alpha/2$ and repeat step 6

Some clarifications are in order. In step 3, we make sure that the determined spectral line step is within certain limits ($[\sigma_{\min}, \sigma_{\max}] = [10^{-3}, 10^3]$). While in our tests the line

step seems to be quite stable, and far away from the limits, one cannot guarantee that this will always be the case.

In step 5, the initial value of the backtracking parameter (α_0) is usually set to 1 in literature [113, 119]. Lastly, step 6 represents the backtracking procedure, inspired by the non-monotone line search technique of Grippo et al. [117]. This checks if the current line-step not only ensures that the residual decreases ($f(\mathbf{X}_k^{\theta\phi} - \lambda_k \alpha \mathbf{Z}_k^{\theta\phi}) < f(\mathbf{X}_k^{\theta\phi})$), but that it decreases by more than $|\gamma\alpha\lambda_k \langle \nabla f(\mathbf{X}_k^{\theta\phi}), \mathbf{Z}\rangle|$, where we found the optimal value of the dampening parameter γ to be $\gamma = 10^{-4}$. This condition is also called the sufficient-decrease condition.

4.7 Validation tests

In order to verify the validity of the implementation, we have utilised our code to simulate the vibrational properties of molecules and crystals.

For the molecular system, we have chosen a simple ethene dimer (C_2H_4), which was simulated using LDA in a cubic cell of 30.0 bohr lateral size, and which was optimised until forces were smaller than 10^{-4} a.u/bohr ($\approx 5 \cdot 10^{-3}$ eV/Å). We have employed both norm-conserving (NC) and projector-augmented-wave (PAW) potentials (Garrity et al. [70]), a 800 eV kinetic energy cutoff, NGWFs of 10.0 bohr radius (one for each H, four for each C).

First, we tested the behaviour of our novel response kernel optimisation (section 4.6). For this, we observed how the response kernel convergence (i.e. decrease in residual $\sqrt{f(\mathbf{X}^{\theta\phi})}$ from Eq. 4.6.3) is affected by the number of preconditioning iterations. In the left inset of Fig. 4.1 one can easily observe the linear convergence behaviour of the quasi-Newton algorithm, enabled by our preconditioning scheme. The number of preconditioning steps seems to be more important for smaller residuals, but as little as 10 preconditioning steps are enough to ensure a well-defined behaviour. The right inset of Fig. 4.1 shows the conjugate gradient (CG) algorithm used for the optimisation of the first-order NGWFs. Due to the much larger variational freedom, it first exhibits linear behaviour but quickly becomes sublinear near convergence. The reason for this mildly suboptimal behaviour

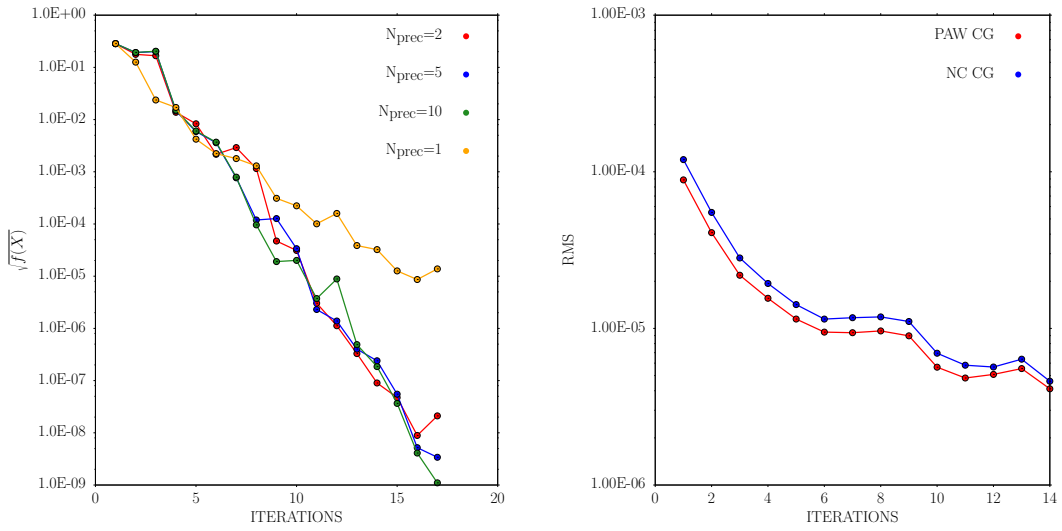


Figure 4.1: Left inset: residual minimisation for quasi-Newton algorithm used in the response kernel optimisation, as a function of the number of preconditioning loops. Right inset: root mean square (RMS) minimisation for conjugate-gradient algorithm used in first-order NGWF optimisation, for PAW and NC pseudopotentials.

was explained at the end of Sec. 4.4.6.

Concerning the vibrational frequencies, in Table 4.1 we have compared our results with an equivalent norm-conserving linear-response phonon calculation performed with the CASTEP plane-wave code [120, 121]. We also include a ONETEP finite-difference phonon calculation, with displacements of 0.1 bohr. We have not included rotational and vibrational frequencies (six in total for molecules), since they have been projected out through the acoustic sum rules for the dynamical matrix

$$\sum_{k'} D^{k\alpha,k'\beta}(\mathbf{q}=\mathbf{0}) = 0 \tag{4.7.1}$$

and the Born charges

$$\sum_{k} Z^{B}_{k\alpha\beta} = 0 \tag{4.7.2}$$

where k, k' are atoms, and α, β are perturbation directions. These rules ensure that the vibrational properties of a system are not affected by global translations or possibly rotations (depending on system dimensionality). We also show a graphical representation of the phonons density of states in Fig. 4.2. It is clear that our linear-response results for both NC and PAW cases match very well both the finite-differencing benchmark of ONETEP, as well as the CASTEP linear-response results. Lastly, in table 4.2, we can

Table 4.1: Vibrational frequencies (in cm^{-1}) for the ethene dimer, obtained from CASTEP linear-response, ONETEP finite-difference (FD), and ONETEP linear-response (LR) phonons using NC and PAW pseudopotentials. Acoustic mode frequencies (corresponding to translation and rotations) are projected out through acoustic sum rules. All units are in cm^{-1}.

Mode	CASTEP NC	ONETEP FD NC	ONETEP LR NC	ONETEP LR PAW
7	804.52	824.74	816.75	807.27
8	932.01	946.47	939.37	931.11
9	946.18	950.71	943.57	938.89
10	1033.54	1050.53	1033.60	1030.05
11	1192.40	1195.70	1190.77	1180.30
12	1325.99	1332.11	1329.54	1320.10
13	1403.29	1404.69	1405.63	1394.27
14	1641.93	1648.43	1639.42	1647.73
15	3027.28	3039.51	3034.51	3050.77
16	3040.48	3049.48	3047.61	3064.28
17	3106.44	3118.41	3113.70	3129.45
18	3133.87	3147.39	3140.78	3153.97

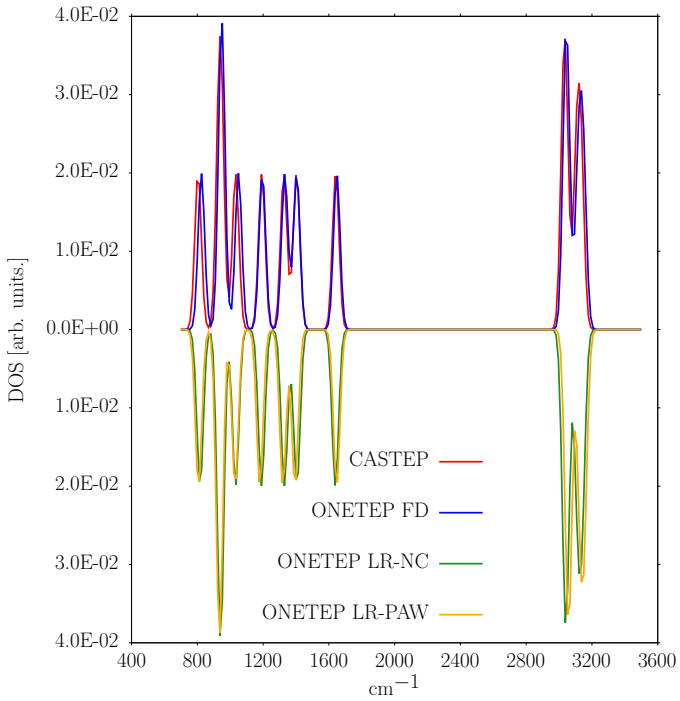


Figure 4.2: Vibrational density of states for ethene dimer, using a broadening of 20 cm^{-1}.

show a direct comparison between the Born charges (Sec. 4.5) obtained with CASTEP and those obtained by our linear-response calculations - for brevity, only the PAW results are shown. Clearly, the agreement is once again very good, proving once again that our code safe to use in molecular systems. Moving onto the periodic crystal system, we have

Table 4.2: Born charges (in é) for the ethene dimer, obtained from CASTEP linear-response and ONETEP linear-response (LR) phonons with PAW pseudopotentials.

Mode	CASTEP			ONETEP		
1	0.0235	-0.0251	-0.0251	0.0240	-0.0266	-0.0255
2	-0.0607	0.0265	-0.0891	-0.0583	0.0253	-0.0868
3	-0.0494	-0.0870	0.0888	-0.0472	-0.0852	0.0860
4	0.0242	0.0321	0.0142	0.0246	0.0330	0.0154
5	0.0677	0.0207	-0.0852	0.0651	0.0198	-0.0831
6	0.0385	-0.0874	0.0938	0.0365	-0.0855	0.0917
7	0.0242	0.0321	0.0142	0.0246	0.0330	0.0154
8	0.0677	0.0207	-0.0852	0.0651	0.0198	-0.0831
9	0.0385	-0.0874	0.0938	0.0365	-0.0855	0.0917
10	0.0235	-0.0251	-0.0251	0.0239	-0.0266	-0.0254
11	-0.0607	0.0265	-0.0891	-0.0583	0.0253	-0.0868
12	-0.0494	-0.0870	0.0888	-0.0472	-0.0852	0.0860
13	-0.0476	-0.0070	0.0109	-0.0485	-0.0064	0.0101
14	-0.0070	-0.0472	0.1743	-0.0068	-0.0451	0.1699
15	0.0109	0.1744	-0.1826	0.0107	0.1707	-0.1777
16	-0.0476	-0.0070	0.0109	-0.0486	-0.0064	0.0101
17	-0.0070	-0.0472	0.1743	-0.0068	-0.0451	0.1699
18	0.0109	0.1744	-0.1826	0.0108	0.1707	-0.1777

used hexagonal boron nitride (hBN) as test subject, in order to verify the usefulness of the code in the context of two-dimensional materials. Our cell needed to be larger than the NGWF diameter; since hBN is well behaved even for small NGWF radii (Sec. 3.1.2), we have utilised 8.0 bohr NGWFs and a 4×4 supercell. Otherwise, we have used the same parameters as for the ethene dimer, with the exception that now the cell has a height of 10.0 Å, which is sufficient for avoiding interactions with periodic images. The comparison between the vibrational frequencies obtained with different methods is shown in Fig. 4.3. It is clear that the frequencies obtained from ONETEP-FD and CASTEP calculations are displaced by a few tens of cm^{-1} due to the inherently non-linear terms present in

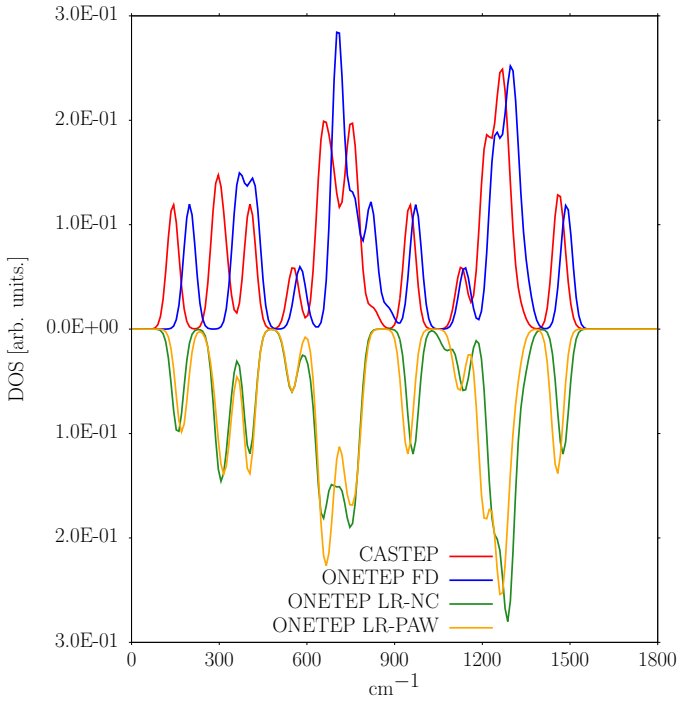


Figure 4.3: Vibrational density of states for a 4×4 hBN supercell, using a broadening of 20 cm^{-1}.

the finite-difference calculation. Moreover, some well defined peaks from the CASTEP results become split in the ONETEP-FD ones. This is due to the presence of symmetry-equivalent atoms in the former code, while the latter does not yet have the capability for symmetry analysis. However, ONETEP-FD does identify a three-fold degeneracy in the vibrational frequencies of hBN. The linear-response ONETEP calculations much better with CASTEP than the ONETEP-FD ones, and our linear-response calculation also identifies the three-fold degeneracy observes in hBN without the need of explicit symmetry-equivalent atoms. We thus conclude that our code is also suitable for periodic crystals at $\mathbf{q} = \mathbf{0}$.

Lastly, we note that we have indeed implemented the framework for non-zero phonon $\mathbf{q}$-wavevectors, electric field perturbations, and acoustic-sum rule enforcement. However, these capabilities still require extensive testing and debugging, the scope of which is beyond this thesis.

Chapter 5

Conclusions

5.1 Two-dimensional heterostructures

As a conclusion to Chapter 3, we have shown that hBN encapsulation does not worsen the desirable electronic properties of pristine monolayer BP. This offers the prospect of stable passivated flakes, allowing BP devices to function in ambient conditions [38, 35]. Moreover, hBN spacers can be used to counteract the band gap reduction in stacked BP, thus improving the efficiency of BP optoelectronic devices in the near-IR range, and allowing more control over light absorption frequencies. Furthermore, based on the tunability of the BP/hBN/BP band gap under electric fields, we have proposed and simulated a tunnelling field effect transistor built from such spaced BP bilayers. Lastly, different operational modes were identified for this transistor, with applications ranging from ultra-fast switching to low-power operation, or enormous peak-to-valley ratios as part of negative differential resistance regimes.

We have also studied TMDC heterostructures, the use of which is more widespread throughout the 2D-material research community. Our work has proven that LS-DFT applied to such interfaces can reliably account for bandstructure effects observed with experimental ARPES [44]. Moreover, our results explicitly show that commensurate and incommensurate domains coexist in $MoSe_2/WSe_2$ heterostructures with low-rotation angles, effects which can be generalised to other TMDC pairs.

I believe that simulating 2D vertical heterostructures should remain a priority for future research; there is no danger of running out of candidates, considering that there are more than 500 layered materials [122], most of which are unexplored. Moreover, the 2D community is aware that stacking different layers can combine their qualities, so the field of van-der-Waals heterostructures is unlikely to become outdated in the future.

5.2 Linear-response phonons

To summarise the second part of the report, we have derived the theoretical details for a linear-response phonon routine for insulators, using the PAW formalism, in the LS-DFT framework of the ONETEP code. Due to ONETEP using a basis of localised NGWFs, we were required to consider innovative methods for optimising the first-order wavefunction, and for displacing the atoms. While linear-response phonons using PAW has already been implemented in plane-wave codes, to the best of our knowledge this is the first instance of such a framework being implemented for localised atomic orbitals, in a linear-scaling manner.

Even if we have set up an initial working system for linear-response phonons, many improvements can still be made in terms of algorithm optimisation, finalising developments and tests for non-zero wave-vectors and electric-field perturbations, and extension to third-order energy terms required for Raman spectroscopy. However, the latter terms would still only require the self-consistent first-order wavefunction, the formalism for which we have already developed and tested. Therefore, the ONETEP community has a functional platform to which it can add the remaining non-selfconsistent components necessary for linear-scaling Raman spectroscopy, of crucial interest to the 2D community.

Appendices

Appendix A

Variational second-order energy

We split the total contribution into five terms ($\frac{d^2 E_{TOT}^{EL-i}}{d\lambda_{-\mathbf{q}} d\epsilon_{\mathbf{q}}}\Big|_{\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}=0}$, with $i = 1,5$), and calculate them separately. The first one, dealing with kinetic and local pseudopotential terms, is:

$$\frac{d^2 E_{\mathrm{TOT}}^{\mathrm{EL}-1}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = \sum_i^{N_e/2} f_i \Bigg\{ \langle\widetilde{\psi}_i\Big|\frac{\partial^2 v_H[\widetilde{\rho}_{Z_c}]}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}\Big|\frac{\partial v_H[\widetilde{\rho}_{Z_c}]}{\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i\Big|\frac{\partial v_H[\widetilde{\rho}_{Z_c}]}{\partial\epsilon_{\mathbf{q}}}\Big]\Big|\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}\rangle +$$
$$\langle\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}\Big|\frac{\partial v_H[\widetilde{\rho}_{Z_c}]}{\partial\lambda_{-\mathbf{q}}}\Big]\Big|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i\Big|\frac{\partial v_H[\widetilde{\rho}_{Z_c}]}{\partial\lambda_{-\mathbf{q}}}\Big]\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}\Big|\Big[-\frac{1}{2}\nabla^2 + v_H[\widetilde{\rho}_{Z_c}]\Big]\Big|\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}\rangle +$$
$$\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}\Big|\Big[-\frac{1}{2}\nabla^2 + v_H[\widetilde{\rho}_{Z_c}]\Big]\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle\Bigg\} = 2\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}\Big|\Big[-\frac{1}{2}\nabla^2 + v_H[\widetilde{\rho}_{Z_c}]\Big]\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle\Bigg\} + \tag{A.0.1}$$
$$\sum_i^{N_e/2} f_i \Bigg\{ \langle\widetilde{\psi}_i\Big|\frac{\partial^2 v_H[\widetilde{\rho}_{Z_c}]}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + 2\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}\Big|\frac{\partial v_H[\widetilde{\rho}_{Z_c}]}{\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + 2\,\langle\widetilde{\psi}_i\Big|\frac{\partial v_H[\widetilde{\rho}_{Z_c}]}{\partial\lambda_{-\mathbf{q}}}\Big]\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle$$

The second term acts as a constraint for the orthogonality between wavefunctions of different bands and k-points:

$$\frac{d^2 E_{TOT}^{EL-2}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = -\frac{d^2}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} \sum_{i,j=1}^{N_e/2}\sum_{\mathbf{k}} f_i \Lambda_{ij\mathbf{k}}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big[\langle\widetilde{\psi}_i[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]|\hat{S}|\widetilde{\psi}_{j,\mathbf{k}}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\rangle - \delta_{ij}\delta_{\mathbf{0},\mathbf{k}}\Big] =$$
$$-\frac{d^2}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} \sum_{i,j=1}^{N_e/2}\sum_{\mathbf{k}} f_i \Bigg[\Lambda_{ij\mathbf{k}}[\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}}]\Big(\langle\widetilde{\psi}_i| + \lambda_{\mathbf{q}}\langle\widetilde{\psi}_i^{1\lambda_{-\mathbf{q}}}| + \lambda_{-\mathbf{q}}\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}| + \langle\epsilon_{-\mathbf{q}}\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}| + \langle\epsilon_{\mathbf{q}}\widetilde{\psi}_i^{1\epsilon_{-\mathbf{q}}}|\Big)$$
$$\hat{S}\Big(|\widetilde{\psi}_{j,\mathbf{k}}\rangle + \lambda_{-\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{k}}^{1\lambda_{-\mathbf{q}}}\rangle + \lambda_{\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{k}}^{1\lambda_{\mathbf{q}}}\rangle + \epsilon_{\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{\mathbf{q}}}\rangle + \epsilon_{-\mathbf{q}}|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{-\mathbf{q}}}\rangle\Big) - \delta_{ij}\delta_{\mathbf{0},\mathbf{k}}\Bigg] = \tag{A.0.2}$$
$$-\sum_{i,j=1}^{N_e/2}\sum_{\mathbf{k}} f_i \Bigg\{\Bigg[\Lambda_{ij\mathbf{k}}^{1\lambda_{-\mathbf{q}}}\Big(\langle\widetilde{\psi}_{j,-\mathbf{k}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_{j,-\mathbf{k}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{\mathbf{q}}}\rangle$$
$$\frac{d^2\Lambda_{ij\mathbf{k}}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}}\Big(\langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,\mathbf{k}}\rangle - \delta_{ij}\delta_{\mathbf{0},\mathbf{k}}\Big) + \Lambda_{ij\mathbf{k}}^{1\epsilon_{\mathbf{q}}}\Big(\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{k}}\rangle + \langle\widetilde{\psi}_{j,-\mathbf{k}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{k}}\rangle\Big)\Big)\Bigg] +$$
$$+\Lambda_{ij\mathbf{k}}\Bigg[\langle\widetilde{\psi}_{j,-\mathbf{k}}|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_{j,-\mathbf{k}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{k}}\rangle + \langle\widetilde{\psi}_{j,-\mathbf{k}}^{1\lambda_{\mathbf{q}}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle +$$
$$\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{k}}^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{k}}\rangle\Bigg]\Bigg\}$$

In the last equality, we have rearranged some terms by using the time-reversal symmetry of wavefunctions ($\langle \mathbf{r}|\widetilde{\psi}_{j\mathbf{k}}\rangle = \langle \widetilde{\psi}_{j-\mathbf{k}}|\mathbf{r}\rangle$). The red term is exactly the wavefunction orthonormality condition, meaning that it reduces to zero. Moreover, it can be shown that the overlap between wavefunctions (both ground-state and perturbed) at different k-points imposes certain selection rules for $\mathbf{k}$ to be either 0, $\mathbf{q}$ or $-\mathbf{q}$ in this case. Therefore, we can finally write the second-order energy term as:

$$\frac{d^2E_{TOT}^{EL-2}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = -\sum_{i,j=1}^{N_e/2} f_i \Bigg\{ \Lambda_{ij\mathbf{q}}^{1\epsilon_{\mathbf{q}}} \Big(\langle \widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle + \langle \widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle + \langle \widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle \Big) + \tag{A.0.3}$$

$$\Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}} \Big(\langle \widetilde{\psi}_{j,\mathbf{q}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle \widetilde{\psi}_{j,\mathbf{q}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle \widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,-\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\rangle \Big) + \Lambda_{ij} \Bigg[\langle \widetilde{\psi}_j|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle \widetilde{\psi}_j^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle +$$

$$\langle \widetilde{\psi}_i^{\lambda_{\mathbf{q}}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_j\rangle + \langle \widetilde{\psi}_j^{1\lambda_{\mathbf{q}}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle \widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_j^{1\epsilon_{\mathbf{q}}}\rangle + \langle \widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_j^{1\epsilon_{\mathbf{q}}}\rangle + \langle \widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_j\rangle \Bigg] \Bigg\}$$

By remembering that $\Lambda_{ij} = \epsilon_i \delta_{ij}$, we obtain:

$$\frac{d^2E_{TOT}^{EL-2}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = -\sum_{i,j=1}^{N_e/2} f_i \Bigg[\Lambda_{ij\mathbf{q}}^{1\epsilon_{\mathbf{q}}} \Big(\langle \widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle + \langle \widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle + \langle \widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle \Big) + \tag{A.0.4}$$

$$\Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}} \Big(\langle \widetilde{\psi}_{j,\mathbf{q}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle \widetilde{\psi}_{j,\mathbf{q}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle \widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,-\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\rangle \Big) \Bigg] -$$

$$\sum_i^{N_e/2} f_i\epsilon_i \Big[2\,\langle \widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + 2\,\langle \widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + 2\,\langle \widetilde{\psi}_i^{\lambda_{\mathbf{q}}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle \widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_j\rangle \Big] \Bigg\}$$

Moving on to the third term, we can simplify expressions by omitting perturbed terms on the atom-centered grid, an approximation which we have previously described in the main text. Thus, we obtain:

$$\frac{d^2E_{TOT}^{EL-3}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = \frac{d}{d\lambda_{-\mathbf{q}}} \sum_{a,b} \Bigg\{ \frac{d\rho^{ab}}{d\epsilon_{\mathbf{q}}} \Bigg[\int v_H\big[\widetilde{\rho}_{Z_c}\big](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\, \mathrm{d}\mathbf{r} + \langle \varphi_a \Big| -\frac{1}{2}\nabla^2 + v_H[\rho_{Z_c}] \Big| \varphi_b\rangle -$$

$$\langle \widetilde{\varphi}_a \Big| -\frac{1}{2}\nabla^2 + v_H[\widetilde{\rho}_{Z_c}] \Big| \widetilde{\varphi}_b\rangle - \int_\Omega v_H[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\, \mathrm{d}\mathbf{r} \Bigg] + \rho^{ab} \Bigg[\int v_H^{1\epsilon_{\mathbf{q}}}[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\, \mathrm{d}\mathbf{r} +$$

$$\int v_H\Big[\widetilde{\rho}_{Z_c}\Big](\mathbf{r})\hat{Q}_{ab}^{1\epsilon_{\mathbf{q}}}(\mathbf{r})\, \mathrm{d}\mathbf{r} \Bigg] \Bigg\} \tag{A.0.5}$$

By explicitly performing the second differentiation, one obtains:

$$\frac{d^2E_{TOT}^{EL-3}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} = \sum_{a,b} \Bigg\{ \frac{d^2\rho^{ab}}{d\lambda_{-\mathbf{q}}\, d\epsilon_{\mathbf{q}}} \Bigg[\int v_H[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\, \mathrm{d}\mathbf{r} + \langle \varphi_a \Big| -\frac{1}{2}\nabla^2 + v_H[\rho_{Z_c}] \Big| \varphi_b\rangle -$$

$$\langle\widetilde{\varphi}_a\Big|-\frac{1}{2}\nabla^2+v_H[\widetilde{\rho}_{Z_c}]\Big|\widetilde{\varphi}_b\rangle-\int_\Omega v_H[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r}\Bigg]+$$
$$\frac{d\rho^{ab}}{d\lambda_{-\mathbf{q}}}\Bigg[\int v_H^{1\epsilon_\mathbf{q}}[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r}+\int v_H\big[\widetilde{\rho}_{Z_c}\big](\mathbf{r})\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}(\mathbf{r})\ \mathrm{d}\mathbf{r}\Bigg]+$$
$$\frac{d\rho^{ab}}{d\epsilon_\mathbf{q}}\Bigg[\int v_H^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}(\mathbf{r})\ \mathrm{d}\mathbf{r}+\int v_H\big[\widetilde{\rho}_{Z_c}\big](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})\ \mathrm{d}\mathbf{r}\Bigg]+ \tag{A.0.6}$$
$$\rho^{ab}\Bigg[\int\frac{\partial^2 v_H[\widetilde{\rho}_{Z_c}]}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_\mathbf{q}}(\mathbf{r})\hat{Q}_{ab}(\mathbf{r})+\int v_H^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}(\mathbf{r})+$$
$$\int v_H^{1\epsilon_\mathbf{q}}[\widetilde{\rho}_{Z_c}](\mathbf{r})\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})+\int v_H[\widetilde{\rho}_{Z_c}](\mathbf{r})\frac{\partial^2\hat{Q}_{ab}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_\mathbf{q}}(\mathbf{r})\Bigg]\Bigg\}$$

Now, we obtain the fourth term in the second-order electronic energy:

$$\frac{d^2E_{TOT}^{EL-4}}{d\lambda_{-\mathbf{q}}\ d\epsilon_\mathbf{q}}=\frac{d}{d\lambda_{-\mathbf{q}}}\Bigg\{\int v_H[\widetilde{\rho}+\hat{\rho}](\mathbf{r})\Big(\widetilde{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})+\hat{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})\Big)\ \mathrm{d}\mathbf{r}+$$
$$\int v_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\Big(\widetilde{\rho}^{1\epsilon_\mathbf{r}}(\mathbf{r})+\hat{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})+\widetilde{\rho}_c^{1\epsilon_\mathbf{q}}(\mathbf{r})\Big)\ \mathrm{d}\mathbf{r}+$$
$$\int_\Omega v_H[\rho^1](\mathbf{r})\rho^{1\ 1\epsilon_\mathbf{q}}(\mathbf{r})\ \mathrm{d}\mathbf{r}+\int_\Omega v_{\mathrm{XC}}[\rho^1+\rho_c](\mathbf{r})\Big(\rho^{1\ 1\epsilon_\mathbf{q}}(\mathbf{r})\Big)\ \mathrm{d}\mathbf{r}- \tag{A.0.7}$$
$$\int_\Omega v_H[\widetilde{\rho}^1+\hat{\rho}](\mathbf{r})\Big(\widetilde{\rho}^{1\ 1\epsilon_\mathbf{q}}(\mathbf{r})+\hat{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})\Big)\ \mathrm{d}\mathbf{r}-\int_\Omega v_{\mathrm{XC}}[\widetilde{\rho}^1+\hat{\rho}+\widetilde{\rho}_c]\Big(\widetilde{\rho}^{1\ 1\epsilon_\mathbf{q}}(\mathbf{r})+\hat{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})\Big)\Bigg\}$$

Upon the explicit differentiation of Eq. A.0.7, and by expanding outside the potentials the augmentation densities (Eq. 4.8) and the compensation density (Eq. 4.9) one obtains:

$$\frac{d^2E_{TOT}^{EL-4}}{d\lambda_{-\mathbf{q}}\ d\epsilon_\mathbf{q}}=\int v_H^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}](\mathbf{r})\Big(\widetilde{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})+\sum_{a,b}\rho^{ab}\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}(\mathbf{r})+\sum_{a,b}\rho^{ab\ 1\epsilon_\mathbf{q}}\hat{Q}_{ab}(\mathbf{r})\Big)\ \mathrm{d}\mathbf{r}+$$
$$\int v_H[\widetilde{\rho}+\hat{\rho}](\mathbf{r})\Bigg(\frac{d^2\widetilde{\rho}}{d\lambda_{-\mathbf{q}}d\epsilon_\mathbf{q}}(\mathbf{r})+\sum_{a,b}\frac{d^2\rho^{ab}}{d\lambda_{-\mathbf{q}}d\epsilon_\mathbf{q}}\hat{Q}_{ab}(\mathbf{r})+\sum_{a,b}\rho^{ab}\frac{\partial^2Q_{ab}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_\mathbf{q}}(\mathbf{r})+ \tag{A.0.8}$$
$$\sum_{a,b}\rho^{ab\ 1\epsilon_\mathbf{q}}\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})+\sum_{a,b}\rho^{ab\ 1\lambda_{-\mathbf{q}}}\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}(\mathbf{r})\Bigg)\ \mathrm{d}\mathbf{r}+\int v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\Bigg(\widetilde{\rho}^{1\epsilon_\mathbf{q}}(\mathbf{r})+$$
$$\sum_{a,b}\rho^{ab}\partial\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}(\mathbf{r})+\sum_{a,b}\rho^{ab\ 1\epsilon_\mathbf{q}}\hat{Q}_{ab}(\mathbf{r})+\widetilde{\rho}_c^{1\epsilon_\mathbf{q}}(\mathbf{r})\Big)\ \mathrm{d}\mathbf{r}+\int v_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\Bigg(\frac{d^2\widetilde{\rho}}{d\lambda_{-\mathbf{q}}d\epsilon_\mathbf{q}}(\mathbf{r})+$$
$$\sum_{a,b}\frac{d^2\rho^{ab}}{d\lambda_{-\mathbf{q}}d\epsilon_\mathbf{q}}\hat{Q}_{ab}(\mathbf{r})+\sum_{a,b}\rho^{ab}\frac{\partial^2Q_{ab}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_\mathbf{q}}(\mathbf{r})+\sum_{a,b}\rho^{ab\ 1\epsilon_\mathbf{q}}\hat{Q}_{ab}^{1\lambda_{-\mathbf{q}}}(\mathbf{r})+\sum_{a,b}\rho^{ab\ 1\lambda_{-\mathbf{q}}}\hat{Q}_{ab}^{1\epsilon_\mathbf{q}}(\mathbf{r})+$$
$$\frac{\partial^2\widetilde{\rho}_c}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_\mathbf{q}}(\mathbf{r})\Bigg)\ \mathrm{d}\mathbf{r}+\int_\Omega\Bigg(v_H^{1\lambda_{-\mathbf{q}}}[\rho^1](\mathbf{r})+v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\rho^1+\rho_c](\mathbf{r})\Bigg)\sum_{a,b}\rho^{ab\ 1\epsilon_\mathbf{q}}\varphi_a(\mathbf{r})\varphi_b(\mathbf{r})\ \mathrm{d}\mathbf{r}+$$

$$\int_\Omega \left(v_H[\rho^1](\mathbf{r}) + v_{\mathrm{XC}}[\rho^1 + \rho_c](\mathbf{r}) \right) \sum_{a,b} \frac{d^2\rho^{ab}}{d\lambda_{-\mathbf{q}} d\epsilon_{\mathbf{q}}} \varphi_a(\mathbf{r})\varphi_b(\mathbf{r})\ \mathrm{d}\mathbf{r} -$$
$$\int_\Omega \sum_{a,b} \rho^{ab\ 1\epsilon_{\mathbf{q}}} \left(\widetilde{\varphi}_a(\mathbf{r})\widetilde{\varphi}_b(\mathbf{r}) + \hat{Q}_{ab}(\mathbf{r}) \right) \left(v_H^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}^1 + \hat{\rho}](\mathbf{r}) + v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}^1 + \hat{\rho} + \widetilde{\rho}_c](\mathbf{r}) \right) \mathrm{d}\mathbf{r} -$$
$$\int_\Omega \left(v_H[\widetilde{\rho}^1 + \hat{\rho}](\mathbf{r}) + v_{XC}[\widetilde{\rho}^1 + \hat{\rho} + \widetilde{\rho}_c](\mathbf{r}) \right) \sum_{a,b} \frac{d^2\widetilde{\rho}^{ab}}{d\lambda_{-\mathbf{q}} d\epsilon_{\mathbf{q}}} \left(\widetilde{\varphi}_a(\mathbf{r})\widetilde{\varphi}_b(\mathbf{r}) + \hat{Q}_{ab}(\mathbf{r}) \right) \mathrm{d}\mathbf{r}$$

As for the last term in the second order energy, it is related to the interaction with external electric fields, if any is present. Moreover, it is only considered for the Gamma point ($\mathbf{q} = \mathbf{0}$) case. Thus, one has:

$$\frac{d^2 E_{TOT}^{EL-5}}{d\lambda_{\mathbf{0}} d\epsilon_{\mathbf{0}}} \delta_{\mathbf{q},\mathbf{0}} = \delta_{\mathbf{q},\mathbf{0}} \frac{d^2}{d\lambda_{\mathbf{0}} \epsilon_{\mathbf{0}}} \sum_{i=1}^{N_e/2} f_i \Big(- \overrightarrow{E} \cdot \langle \widetilde{\psi}_i | (-\overrightarrow{\hat{\mathcal{D}}}) | \widetilde{\psi}_i \rangle \Big) = \delta_{\mathbf{q},\mathbf{0}} \frac{d}{d\lambda_{\mathbf{0}}} \sum_{i=1}^{N_e/2} f_i \Bigg[$$
$$\frac{\partial \overrightarrow{E}}{\partial \epsilon_{\mathbf{0}}} \cdot \langle \widetilde{\psi}_i | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i \rangle + \overrightarrow{E} \cdot \langle \widetilde{\psi}_i^{1\epsilon_{\mathbf{0}}} | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i \rangle + \overrightarrow{E} \cdot \langle \widetilde{\psi}_i | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i^{1\epsilon_{\mathbf{0}}} \rangle + \overrightarrow{E} \cdot \langle \widetilde{\psi}_i | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\epsilon_{\mathbf{0}}} | \widetilde{\psi}_i \rangle \Bigg] = \tag{A.0.9}$$
$$\sum_{i=1}^{N_e/2} f_i \frac{\partial \overrightarrow{E}}{\partial \epsilon_{\mathbf{0}}} \cdot \left[\langle \widetilde{\psi}_i^{1\lambda_{\mathbf{0}}} | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\lambda_{\mathbf{0}}} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i^{1\lambda_{\mathbf{0}}} \rangle \right] \delta_{\mathbf{q},\mathbf{0}} +$$
$$\sum_{i=1}^{N_e/2} f_i \frac{\partial \overrightarrow{E}}{\partial \lambda_{\mathbf{0}}} \cdot \left[\langle \widetilde{\psi}_i^{1\epsilon_{\mathbf{0}}} | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\epsilon_{\mathbf{0}}} | \widetilde{\psi}_i \rangle + \langle \widetilde{\psi}_i | \overrightarrow{\mathcal{D}} | \widetilde{\psi}_i^{1\epsilon_{\mathbf{0}}} \rangle \right] \delta_{\mathbf{q},\mathbf{0}} +$$
$$\sum_{i=1}^{N_e/2} f_i \overrightarrow{E} \cdot \Bigg[\langle \psi_i^{1\lambda_{\mathbf{0}}} | \overrightarrow{\hat{\mathcal{D}}} | \psi_i^{1\epsilon_{\mathbf{0}}} \rangle + \langle \psi_i^{1\epsilon_{\mathbf{0}}} | \overrightarrow{\hat{\mathcal{D}}} | \psi_i^{1\lambda_{\mathbf{0}}} \rangle + \langle \psi_i^{1\lambda_{\mathbf{0}}} | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\epsilon_{\mathbf{0}}} | \psi_i \rangle + \langle \psi_i | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\epsilon_{\mathbf{0}}} | \psi_i^{1\lambda_{\mathbf{0}}} \rangle +$$
$$\langle \psi_i^{1\epsilon_{\mathbf{0}}} | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\lambda_{\mathbf{0}}} | \psi_i \rangle + \langle \psi_i | \frac{d\overrightarrow{\hat{\mathcal{D}}}}{d\lambda_{\mathbf{0}}} | \psi_i^{1\epsilon_{\mathbf{0}}} \rangle + \langle \psi_i | \frac{d^2\overrightarrow{\hat{\mathcal{D}}}}{d\lambda_{\mathbf{0}} d\epsilon_{\mathbf{0}}} | \psi_i \rangle \Bigg] \delta_{\mathbf{q},\mathbf{0}} + \sum_{i=1}^{N_e/2} f_i \frac{\partial^2 \overrightarrow{E}}{\partial \lambda_{\mathbf{0}} \partial \epsilon_{\mathbf{0}}} \cdot \langle \psi_i | \overrightarrow{\hat{\mathcal{D}}} | \psi_i \rangle\ \delta_{\mathbf{q},\mathbf{0}}$$

Some simplifications can be made to the previous relation, however. Since the presence of an electric field is of interest to us only when we need to obtain the dielectric matrix (i.e. under electric field perturbations), we can safely assume that, for the time being, Eq. A.0.9 is nonzero only when λ and ϵ are electric field perturbations, and not atom displacements. Therefore, since the dielectric matrix does not depend on the electric field, all perturbations of $\overrightarrow{\mathcal{D}}$ are 0 in this case. Similarly, since λ and ϵ are essentially electric field components in this case, it is clear that $\frac{\partial^2 \overrightarrow{E}}{\partial \lambda_{\mathbf{0}} \partial \epsilon_{\mathbf{0}}}$ is also 0.

The final step consists of grouping the terms in Eqs. A.0.1, A.0.4, A.0.6, A.0.8 and A.0.9 in order to obtain the total variational second-order energy. After expanding $\widetilde{\rho}$ and ρ^{ab} and their derivatives (according to Eqs. 4.6 and 4.7), one obtains the following (U is the

Ewald energy, and we have worked in the limit of zero $\overrightarrow{E}$):

$$\frac{d^2E_{TOT}}{d\lambda_{-\mathbf{q}}d\epsilon_{\mathbf{q}}}\Big|_{\lambda_{-\mathbf{q}},\epsilon_{\mathbf{q}},\overrightarrow{E}=0} = \frac{d^2\left(E_{TOT}^{EL-1}+E_{TOT}^{EL-2}+E_{TOT}^{EL-3}+E_{TOT}^{EL-4}+E_{TOT}^{EL-5}\right)}{d\lambda_{-\mathbf{q}}d\epsilon_{\mathbf{q}}} + \frac{\partial^2}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}E_{Ion,core} =$$

$$\sum_{i=1}^{N_e/2} f_i\Bigg\{ \langle\widetilde{\psi}_i\Big|\hat{A}^{\lambda_{-\mathbf{q}}\epsilon_{\mathbf{q}}} - \epsilon_i\frac{\partial^2\hat{S}}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle + 2\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{H}-\epsilon_i\hat{S}\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + 2\,\langle\widetilde{\psi}_i|\frac{d\hat{H}}{d\lambda_{-\mathbf{q}}} - \epsilon_i\frac{\partial\hat{S}}{\partial\lambda_{-\mathbf{q}}}\Big|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle +$$

$$2\,\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{B}^{\epsilon_{\mathbf{q}}} - \epsilon_i\frac{\partial\hat{S}}{\partial\epsilon_{\mathbf{q}}}\Big|\widetilde{\psi}_i\rangle\Bigg\} + \int v_{\mathrm{XC}}^{1\lambda_{-\mathbf{q}}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\widetilde{\rho}_c^{1\epsilon_{\mathbf{q}}}(\mathbf{r})\,\mathrm{d}\mathbf{r} + \tag{A.0.10}$$

$$\int v_{\mathrm{XC}}[\widetilde{\rho}+\hat{\rho}+\widetilde{\rho}_c](\mathbf{r})\frac{\partial^2\widetilde{\rho}_c(\mathbf{r})}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\,\mathrm{d}\mathbf{r} - \sum_{i,j=1}^{N_e/2} f_i\Bigg[\Lambda_{ij\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\Big(\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_{j,\mathbf{q}}\rangle + \langle\widetilde{\psi}_{j,-\mathbf{q}}^{1\lambda_{\mathbf{q}}}|\hat{S}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\hat{S}^{1\lambda_{-\mathbf{q}}}|\widetilde{\psi}_{j,\mathbf{q}}\rangle\Big) +$$

$$\Lambda_{ij-\mathbf{q}}^{1\lambda_{-\mathbf{q}}}\Big(\langle\widetilde{\psi}_{j,\mathbf{q}}|\hat{S}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{q}}}\rangle + \langle\widetilde{\psi}_{j,\mathbf{q}}|\hat{S}^{1\epsilon_{\mathbf{q}}}|\widetilde{\psi}_i\rangle + \langle\widetilde{\psi}_i|\hat{S}|\widetilde{\psi}_{j,-\mathbf{q}}^{1\epsilon_{\mathbf{q}}}\rangle\Big)\Bigg] + \frac{\partial^2}{\partial\lambda_{-\mathbf{q}}\partial\epsilon_{\mathbf{q}}}\left[\frac{1}{2}\int_\Omega\int_\Omega\frac{\rho_{Z_c}(\mathbf{r})\rho_{Z_c}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' + U\right] +$$

$$\sum_{i=1}^{N_e/2} 2f_i\frac{\partial\overrightarrow{E}}{\partial\epsilon_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i^{1\lambda_{\mathbf{0}}}|\overrightarrow{\hat{\mathcal{D}}}|\widetilde{\psi}_i\rangle\,\delta_{\mathbf{q},\mathbf{0}} + \sum_{i=1}^{N_e/2} 2f_i\frac{\partial\overrightarrow{E}}{\partial\lambda_{\mathbf{0}}}\cdot\langle\widetilde{\psi}_i|\overrightarrow{\hat{\mathcal{D}}}|\widetilde{\psi}_i^{1\epsilon_{\mathbf{0}}}\rangle\,\delta_{\mathbf{q},\mathbf{0}}$$

Appendix B

Tunnelling BP/hBN/BP FET

In this appendix we show our complete mathematical formalism for the computation of the tunnelling current and associated coupling matrix terms (Eqs. 3.1.1 and 3.1.2) in BP/hBN interfaces.

B.1 Derivation of the coupling matrix terms

Since our simulation cell contains large black phosphorus supercells, the coupling matrix terms $\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}})$

$$\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}}) = \langle \widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) \mid \mathrm{V} \mid \widetilde{\psi}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{T}}) \rangle \tag{B.1.1}$$

can only be calculated by using an unfolding procedure similar to that employed for the spectral function [36]. Working in the supercell reference frame, we start by inserting two identity operators (in the reference frame of the supercell)

$$\sum_{B=1}^{N} |\phi^{B}{}_{\mathbf{K}}\rangle \langle \phi_{B\mathbf{K}}| \hat{S} = \sum_{B=1}^{N} \left[\sum_{A=1}^{N} |\phi_{A\mathbf{K}}\rangle \ \left(S^{-1}\right)^{AB}(\mathbf{K}) \right] \ \langle \phi_{B\mathbf{K}}| \hat{S} = 1 \tag{B.1.2}$$

on either side of the scattering potential V, at the supercell k-points $\mathbf{K}$ and $\mathbf{K}'$. $\hat{S}$ is the PAW-augmented overlap operator, $\left(S^{-1}\right)^{AB}$ are elements of the representation of its inverse, and $\phi_{A,\mathbf{K}}$ is a complex NGWF with index A (running from 1 to N - the number of NGWFs in the supercell) at the supercell k-point $\mathbf{K}$. Any complex NGWF can be written in terms of the real-space NGWFs $\phi_{A\mathbf{R}}$, where $\mathbf{R}$ is a multiple of the supercell lattice vectors:

$$|\phi_{A\mathbf{K}}\rangle = \frac{1}{\sqrt{L}} \sum_{\mathbf{R}} e^{i\mathbf{K}\cdot\mathbf{R}} \ |\phi_{A\mathbf{R}}\rangle \ , \tag{B.1.3}$$

where L is the number of simulation cell duplicates considered in the sum (over $\mathbf{R}$). Consequently, the identity operator in Eq. B.1.2 becomes:

$$\frac{1}{L} \sum_{A,B=1}^{N} \sum_{\mathbf{R},\mathbf{R}'} e^{i\mathbf{K}\cdot(\mathbf{R}-\mathbf{R}')} \ |\phi_{A,\mathbf{R}}\rangle \ \left(S^{-1}\right)^{AB}(\mathbf{K}) \ \langle \phi_{B,\mathbf{R}'}| \hat{S} = 1 \tag{B.1.4}$$

With the identity operators in place, the transfer matrix term now has the form:

$$\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}},\mathbf{k}_{\mathcal{T}}) = \frac{1}{L^2}\sum_{\mathbf{K},\mathbf{K}'}\sum_{\mathbf{R},\mathbf{R}'}\sum_{\mathbf{R}'',\mathbf{R}'''}\sum_{A,B=1}^{N}\sum_{C,D=1}^{N} e^{i\mathbf{K}\cdot(\mathbf{R}-\mathbf{R}')}\ \langle\widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})|\,\hat{P}^{\dagger}_{\mathcal{B}}\ \hat{S}\,|\phi_{A,\mathbf{R}}\rangle$$
$$\left(S^{-1}\right)^{AB}(\mathbf{K})\ \langle\phi_{B,\mathbf{R}'}|\,V\,|\phi_{C,\mathbf{R}''}\rangle\ \left(S^{-1}\right)^{CD}(\mathbf{K}')\ \langle\phi_{D,\mathbf{R}'''}|\,\hat{S}\hat{P}_{\mathcal{T}}\ |\widetilde{\psi}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{T}})\rangle\, e^{i\mathbf{K}'\cdot(\mathbf{R}''-\mathbf{R}''')} \tag{B.1.5}$$

Since a prerequisite to unfolding a function is to project the latter onto the monolayer of interest[36], in Eq. B.1.5 we were required to place projection operators for the bottom ($\hat{P}^{\dagger}_{\mathcal{B}} = \hat{S}\sum_{\rho\in\mathcal{B}}|\phi^{\rho}\rangle\langle\phi_{\rho}|$) and top monolayers ($\hat{P}_{\mathcal{T}} = \sum_{\eta\in\mathcal{T}}|\phi^{\eta}\rangle\langle\phi_{\eta}|\,\hat{S}$) between the previously discussed identity operators and the primitive-cell eigenvectors of the bottom ($\mathcal{B}$) and top ($\mathcal{T}$) BP layers ($\widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})$ and $\widetilde{\psi}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{T}})$). These wavefunctions can be expanded as

$$|\widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\rangle = \sum_{\alpha=1}^{n}\mathcal{M}^{\alpha}_{\ i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ \ |\phi_{\alpha\mathbf{k}_{\mathcal{B}}}\rangle = \frac{1}{\sqrt{l}}\sum_{\alpha=1}^{n}\mathcal{M}^{\alpha}_{\ i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\sum_{\mathbf{r}} e^{i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}}\ \ |\phi_{\alpha\mathbf{r}}\rangle\ , \tag{B.1.6}$$

where $\mathcal{M}^{\alpha}_{\ i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})$ is the molecular-orbital coefficient of NGWF α with respect to the bottom primitive-cell eigenvector i at the primitive-unit k-point $\mathbf{k}_{\mathcal{B}}$. In Eq. B.1.6, one sums over l duplicates of the primitive-cell, with their position vector $\mathbf{r}$ being a multiple of the primitive cell lattice vectors; also, $n = 36$ is the number of NGWFs in a primitive cell of the black phosphorus monolayer (9 functions for each P atom).

At this point, we need to clarify the notations behind the NGWF indices. The capital letters (A, B, C, D), as well as the Greek letters from the middle of the alphabet (η, μ, ρ, π) correspond to NGWFs of the supercell, i.e. they take values from 1 to N. Letters from the beginning of the Greek alphabet (α, β, α') correspond to NGWFs in the representation of the primitive cell, i.e. they take values from 1 to n.

Consequently, by plugging Eq. B.1.6 into expression for the coupling matrix terms (Eq. B.1.5), one obtains:

$$\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}},\mathbf{k}_{\mathcal{T}}) = \sum_{\substack{\mathbf{K},\mathbf{K}'\\A,B,C,D}}\sum_{\alpha,\beta}\frac{1}{l\ L^2}\mathcal{M}^{\dagger\ \alpha}_{\ i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\mathcal{M}^{\beta}_{\ j\mathcal{T}}(\mathbf{k}_{\mathcal{T}})\ \left(S^{-1}\right)^{AB}(\mathbf{K})\ \left(S^{-1}\right)^{CD}(\mathbf{K}')$$
$$\sum_{\mathbf{r},\mathbf{R}} e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}}\ e^{-i\mathbf{K}\cdot\mathbf{R}}\,\langle\phi_{\alpha\mathbf{r}}|\hat{P}^{\dagger}_{\mathcal{B}}\hat{S}|\phi_{A\mathbf{R}}\rangle\sum_{\mathbf{R}',\mathbf{R}''} e^{-i\mathbf{K}\cdot\mathbf{R}'}\ e^{i\mathbf{K}'\cdot\mathbf{R}''}\,\langle\phi_{B\mathbf{R}'}|V|\phi_{C\mathbf{R}''}\rangle \tag{B.1.7}$$

$$\sum_{\mathbf{r}',\mathbf{R}'''} e^{-i\mathbf{K}'\cdot\mathbf{R}'''}\ e^{i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}''}\ \langle\phi_{D\mathbf{R}'''}|\hat{S}\hat{P}_{\mathcal{T}}|\phi_{\beta\mathbf{r}''}\rangle$$

In the previous expression (Eq. B.1.7), the middle term can be simplified as:

$$\begin{aligned}\sum_{\mathbf{R}',\mathbf{R}''} e^{i[\mathbf{K}'\cdot\mathbf{R}''-\mathbf{K}\cdot\mathbf{R}']}\ \langle\phi_{B\mathbf{R}'}|V|\phi_{C\mathbf{R}''}\rangle &= \sum_{\mathbf{R}',\Delta\mathbf{R}'} e^{i\mathbf{K}'\cdot(\mathbf{R}'+\Delta\mathbf{R}')}\ e^{-i\mathbf{K}\cdot\mathbf{R}'}\ \langle\phi_{B\mathbf{0}}|V|\phi_{C\Delta\mathbf{R}'}\rangle = \\ &= \sum_{\mathbf{R}'} e^{i(\mathbf{K}'-\mathbf{K})\cdot\mathbf{R}'} \sum_{\Delta\mathbf{R}'} e^{i\mathbf{K}'\cdot\Delta\mathbf{R}''}\ \langle\phi_{B\mathbf{0}}|V|\phi_{C\Delta\mathbf{R}'}\rangle = L\delta_{\mathbf{K},\mathbf{K}'}V_{BC}(\mathbf{K}')\end{aligned} \tag{B.1.8}$$

In Eq. B.1.8, for the first equality we have used the translational symmetry of the NGWFs

$$\langle\phi_{B\mathbf{R}'}|V|\phi_{C\mathbf{R}''}\rangle = \langle\phi_{B\mathbf{0}}|V|\phi_{C\mathbf{R}''-\mathbf{R}'}\rangle = \langle\phi_{B\mathbf{0}}|V|\phi_{C\Delta\mathbf{R}'}\rangle\ , \tag{B.1.9}$$

while for the last equality we have utilised the definition:

$$\sum_{\Delta\mathbf{R}'} e^{i\mathbf{K}'\cdot\Delta\mathbf{R}''}\ \langle\phi_{B\mathbf{0}}|V|\phi_{C\Delta\mathbf{R}'}\rangle = V_{BC}(\mathbf{K}') \tag{B.1.10}$$

Now, we focus on the first term of Eq. B.1.5:

$$\sum_{\mathbf{r},\mathbf{R}} e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}}\ e^{i\mathbf{K}\cdot\mathbf{R}}\ \langle\phi_{\alpha\mathbf{r}}|\ \underbrace{\hat{S}\sum_{\rho\in\mathcal{B}}|\phi^{\rho}\rangle\ \langle\phi_{\rho}|}_{\hat{P}^{\dagger}_{\mathcal{B}}}\ \hat{S}\ |\phi_{A\mathbf{R}}\rangle\ , \tag{B.1.11}$$

where $|\phi^{\rho}\rangle$ is the NGWF-dual, defined as:

$$|\phi^{\rho}\rangle = \sum_{\pi}|\phi_{\pi}\rangle\left(S^{-1}\right)^{\pi\rho} \tag{B.1.12}$$

By plugging Eq. B.1.12 into Eq. B.1.11, one obtains:

$$\sum_{\substack{\mathbf{r},\mathbf{R}\\ \rho\in\mathcal{B}}} e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}}\ e^{i\mathbf{K}\cdot\mathbf{R}}\ \underbrace{\sum_{\pi} S_{\alpha\mathbf{r},\pi}\left(S^{-1}\right)^{\pi\rho}}_{\delta_{\alpha\mathbf{r}}{}^{\rho}}\ S_{\rho,A\mathbf{R}} = \sum_{\mathbf{r},\mathbf{R}} e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}}\ e^{i\mathbf{K}\cdot\mathbf{R}}\ S_{\alpha\mathbf{r},A\mathbf{R}}\ , \tag{B.1.13}$$

with the primitive-cell /supercell NGWF indices $\alpha\mathbf{r}$ / $A\mathbf{R}$ now being constrained only to the bottom Black Phosphorus layer. From previous derivations for the spectral function unfolding [36, 123], we know that the cross-overlap between NGWFs in the supercell and primitive-cell representation takes the form:

$$S_{\alpha\mathbf{r},A\mathbf{R}} = \sum_{\mathbf{k}}\frac{1}{l}\ e^{i\mathbf{k}\cdot\left[\mathbf{r}-\mathbf{R}-\mathbf{r}_0(A)\right]}\ S_{\alpha\alpha'(A)}(\mathbf{k}) \tag{B.1.14}$$

Consequently, by replacing Eq. B.1.14 into Eq. B.1.13, we now obtain:

$$\begin{aligned}\sum_{\mathbf{r},\mathbf{R}} e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}}\ e^{i\mathbf{K}\cdot\mathbf{R}} \sum_{\mathbf{k}} \frac{1}{l}\ e^{i\mathbf{k}\cdot\left[\mathbf{r}-\mathbf{R}-\mathbf{r}_0(A)\right]}\ S_{\alpha\alpha'(A)}(\mathbf{k}) = \\ = \sum_{\mathbf{k}} \frac{1}{l}\ e^{-i\mathbf{k}\cdot\mathbf{r}_0(A)} \underbrace{\sum_{\mathbf{r}} e^{i\mathbf{r}\cdot(\mathbf{k}-\mathbf{k}_{\mathcal{B}})}}_{l\delta_{\mathbf{k},\mathbf{k}_{\mathcal{B}}}} \underbrace{\sum_{\mathbf{R}} e^{i\mathbf{R}\cdot(\mathbf{K}-\mathbf{k})}}_{L\sum_{\mathbf{G}}\delta_{\mathbf{k}_{\mathcal{B}}-\mathbf{G},\mathbf{K}}}\ S_{\alpha,\alpha'(A)}(\mathbf{k}) = \\ = L\sum_{\mathbf{G}} e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}_0(A)}\ \delta_{\mathbf{k}_{\mathcal{B}}-\mathbf{G},\mathbf{K}}\ S_{\alpha\alpha'(A)}(\mathbf{k}_{\mathcal{B}})\end{aligned} \tag{B.1.15}$$

Moving up to the last term of Eq. B.1.5, we have:

$$\begin{aligned}\sum_{\mathbf{r}',\mathbf{R}'''} e^{i\mathbf{k}_{\mathcal{T}}\cdot\mathbf{r}'}\ e^{-i\mathbf{K}'\cdot\mathbf{R}'''}\ \langle\phi_{D\mathbf{R}'''}|\ \hat{S}\hat{P}_{\mathcal{T}}\ |\phi_{\beta\mathbf{r}'}\rangle = \sum_{\substack{\mathbf{r}',\mathbf{R}''' \\ \eta\in\mathcal{T}}} e^{i\mathbf{k}_{\mathcal{T}}\cdot\mathbf{r}'}\ e^{-i\mathbf{K}'\cdot\mathbf{R}'''}\ \langle\phi_{D\mathbf{R}'''}| \\ \hat{S}\ |\phi_\eta\rangle \sum_{\mu}\left(S^{-1}\right)^{\eta\mu}\langle\phi_\mu|\hat{S}|\phi_{\beta\mathbf{r}'}\rangle = \sum_{\mathbf{r}',\mathbf{R}'''} e^{i\mathbf{k}_{\mathcal{T}}\cdot\mathbf{r}}\ e^{-i\mathbf{K}'\cdot\mathbf{R}'''}\ S_{D\mathbf{R}''',\beta\mathbf{r}'}\end{aligned} \tag{B.1.16}$$

where $\beta\mathbf{r}'$ and $D\mathbf{R}'''$ are indices for NGWFs corresponding only to the top Black Phosphorus layer. We can further simplify this relation by considering Eqs. B.1.8 and B.1.15, more specifically the terms $\delta_{\mathbf{K},\mathbf{K}'}$ and $\delta_{\mathbf{K},\mathbf{k}_{\mathcal{B}}-\mathbf{G}}$; consequently, Eq. B.1.16 becomes, considering ($\mathbf{k}_{\mathcal{B}} = \mathbf{k}_{\mathcal{T}} + \mathbf{q}$):

$$\begin{aligned}\sum_{\mathbf{r}',\mathbf{R}'''} e^{i\mathbf{k}_{\mathcal{T}}\cdot\mathbf{r}'}\ e^{-i(\mathbf{k}_{\mathcal{B}}-\mathbf{G})\cdot\mathbf{R}'''}\ S_{D\mathbf{R}''',\beta\mathbf{r}'} = \sum_{\mathbf{R}'''}\underbrace{e^{i\mathbf{G}\cdot\mathbf{R}'''}}_{=1}\sum_{\mathbf{r}'} e^{i\mathbf{k}_{\mathcal{T}}\cdot\mathbf{r}'}\ e^{-i(\mathbf{k}_{\mathcal{T}}+\mathbf{q})\cdot\mathbf{R}'''}\ S_{D\mathbf{R}''',\beta\mathbf{r}'} = \\ \sum_{\mathbf{R}'''} e^{-i\mathbf{q}\cdot\mathbf{R}'''}\sum_{\mathbf{r}'} e^{i\mathbf{k}_{\mathcal{T}}\cdot(\mathbf{r}'-\mathbf{R}''')}\ S_{D\mathbf{0},\beta\ \mathbf{r}'-\mathbf{R}'''} = L\ \delta_{\mathbf{q},\mathbf{0}}\ S_{D\beta}(\mathbf{k}_{\mathcal{T}}) = L\ \delta_{\mathbf{q},\mathbf{0}}\ S_{D\beta}(\mathbf{k}_{\mathcal{B}}-\mathbf{q})\end{aligned} \tag{B.1.17}$$

Putting all the terms together, the coupling matrix term becomes:

$$\begin{aligned}\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}},\mathbf{k}_{\mathcal{T}}=\mathbf{k}_{\mathcal{B}}-\mathbf{q}) = \frac{L}{l}\sum_{\mathbf{K},\mathbf{G}}\sum_{B,C=1}^{N}\sum_{\substack{D\in\mathcal{T}\\A\in\mathcal{B}}}\sum_{\alpha,\beta=1}^{n}\mathcal{M}^{\dagger}{}_{i\mathcal{B}}^{\ \alpha}(\mathbf{k}_{\mathcal{B}})\mathcal{M}^{\beta}{}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{B}}-\mathbf{q})\ e^{-i\mathbf{k}_{\mathcal{B}}\cdot\mathbf{r}_0(A)} \\ \delta_{\mathbf{k}_{\mathcal{B}}-\mathbf{G},\mathbf{K}}\ S_{\alpha\alpha'(A)}(\mathbf{k}_{\mathcal{B}})\ \ \left(S^{-1}\right)^{AB}(\mathbf{K})\ V_{BC}(\mathbf{K})\ \left(S^{-1}\right)^{CD}(\mathbf{K})\ \delta_{\mathbf{0},\mathbf{q}}\ S_{D\beta}(\mathbf{k}_{\mathcal{B}}-\mathbf{q})\ ,\end{aligned} \tag{B.1.18}$$

where n is the number of NGWFs present in a primitive cell of Black Phosphorus (i.e. $n = 36$). Further simplifications can be employed in the limit of identical independent layers (i.e. $\widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})$ and $\widetilde{\psi}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{T}})$ reside in independent Hilbert spaces) and of small in-plane momentum variation ($|\mathbf{q}| << |\mathbf{k}_{\mathcal{B}}|$), one can assume:

$$\mathcal{M}^{\beta}{}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{B}}-\mathbf{q}) \approx \mathcal{M}^{\beta}{}_{j\mathcal{B}}(\mathbf{k}_{\mathcal{B}}-\mathbf{q}) \approx \mathcal{M}^{\beta}{}_{j\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) \tag{B.1.19}$$

In the first approximation of Eq. B.1.19, one assumes that the eigenvectors of the different layers have the same corresponding molecular orbital coefficients, differing only through the position of the atoms (i.e. associated NGWFs).

However, we have access only to the supercell representation of the orbital coefficients, but not to the primitive cell ones. Thus, we must transition from one reference frame to another, in several steps.

Firstly, in order to eliminate entirely the primitive-cell orbital coefficients $\mathcal{M}^{\beta}_{\ i\mathcal{B}}$, $\mathcal{M}^{\beta}_{\ j\mathcal{B}}$ from the transfer matrix terms (Eq. B.1.18), we must (temporarily) sum over all state pairs ij. To prove that the aforementioned statement, we first define:

$$X^{\beta\alpha}(\mathbf{k}_{\mathcal{B}}) = \sum_{i,j=1}^{n} \mathcal{M}^{\dagger}{}^{\alpha}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ \mathcal{M}^{\beta}_{\ j\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ , \tag{B.1.20}$$

where n is the total number of eigenstates in the primitive BP monolayer unit. Afterwards, due to the orthonormality of the valence and conduction eigenvectors of the independent bottom BP layer, one has:

$$\begin{aligned} n = \sum_{i,j=1}^{n} \langle \widetilde{\psi}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) | \hat{S} | \widetilde{\psi}_{j\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) \rangle = \sum_{\substack{\alpha,\beta=1,n \\ i,j=1,n}} \frac{1}{l} \mathcal{M}^{\dagger}{}^{\alpha}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ \mathcal{M}^{\beta}_{\ j\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) \sum_{\mathbf{r},\mathbf{r}'} e^{i\mathbf{k}_{\mathcal{B}}\cdot(\mathbf{r}-\mathbf{r}')}\ \langle \phi_{\alpha\mathbf{r}'} | \hat{S} | \phi_{\beta\mathbf{r}} \rangle = \\ \sum_{\alpha,\beta=1}^{n} \sum_{i,j=1}^{n} \frac{1}{l} \mathcal{M}^{\dagger}{}^{\alpha}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ \mathcal{M}^{\beta}_{\ j\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) \underbrace{\Big[\sum_{\mathbf{r},\Delta\mathbf{r}} e^{i\mathbf{k}_{\mathcal{B}}\cdot\Delta\mathbf{r}}\ \langle \phi_{\alpha\mathbf{0}} | \hat{S} | \phi_{\beta\Delta\mathbf{r}} \rangle \Big]}_{=\sum_{\mathbf{r}}\ S_{\alpha\beta}(\mathbf{k}_{\mathcal{B}}) = l S_{\alpha\beta}(\mathbf{k}_{\mathcal{B}})} = \sum_{\alpha,\beta=1}^{n} X^{\beta\alpha}(\mathbf{k}_{\mathcal{B}})\ S_{\alpha\beta}(\mathbf{k}_{\mathcal{B}}) \end{aligned} \tag{B.1.21}$$

where for the underbrace we have used the definition of the k-dependent overlap matrix in the primitive-cell representation:

$$\sum_{\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}}\ \langle \phi_{\alpha\mathbf{0}} | \hat{S} | \phi_{\beta\mathbf{r}} \rangle = S_{\alpha\beta}(\mathbf{k}) \tag{B.1.22}$$

Clearly, the first and last terms in Eq. B.1.21 are equal if

$$\sum_{i,j=1}^{n} \mathcal{M}^{\dagger}{}^{\alpha}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ \mathcal{M}^{\beta}_{\ j\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) = X^{\beta\alpha}(\mathbf{k}_{\mathcal{B}}) = \left(S^{-1}\right)^{\beta\alpha}(\mathbf{k}_{\mathcal{B}}) \tag{B.1.23}$$

Eq. B.1.23 does not come as a surprise, since it is known that the inverse overlap is a sum over all occupied and unoccupied eigenvector products

$$\left(S^{-1}\right)^{\beta\alpha}(\mathbf{k}_{\mathcal{B}}) = \sum_{i=1}^{n} \mathcal{M}^{\dagger}{}^{\alpha}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ \mathcal{M}^{\beta}_{\ i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})\ , \tag{B.1.24}$$

and that the additional sum over j from Eq. B.1.22 has no effect due to the orthogonality between different eigenstates. Thus, by summing over the state pairs ij in Eq. B.1.18 and by utlising Eq. B.1.23, one obtains:

$$\sum_{i,j} \mathrm{O}_{ij}(\mathbf{k}_\mathcal{B}, \mathbf{k}_\mathcal{T} = \mathbf{k}_\mathcal{B} - \mathbf{q}) = \frac{L}{l} \sum_{\mathbf{K},\mathbf{G}} \sum_{D\in\mathcal{T}} \sum_{A\in\mathcal{B}} \overbrace{\sum_{\alpha,\beta} \left(S^{-1}\right)^{\beta\alpha}(\mathbf{k}_\mathcal{B})\ S_{\alpha\alpha'(A)}(\mathbf{k}_\mathcal{B})}^{\sum_\beta \delta^\beta{}_{\alpha'(A)}}$$
$$e^{-i\mathbf{k}_\mathcal{B}\cdot\mathbf{r}_0(A)} \left[\sum_{B,C} \left(S^{-1}\right)^{AB}(\mathbf{K})\ V_{BC}(\mathbf{K})\ \left(S^{-1}\right)^{CD}(\mathbf{K})\right] \delta_{\mathbf{0},\mathbf{q}} \underbrace{\sum_{\mathbf{r}} e^{i(\mathbf{k}_\mathcal{B}-\mathbf{q})\cdot\mathbf{r}}\ S_{D\mathbf{0},\beta\mathbf{r}}}_{S_{D\beta}(\mathbf{k}_\mathcal{B}-\mathbf{q})}\ , \tag{B.1.25}$$

which simplifies to

$$\sum_{i,j=1}^{n} \mathrm{O}_{ij}(\mathbf{k}_\mathcal{B}, \mathbf{k}_\mathcal{T} = \mathbf{k}_\mathcal{B} - \mathbf{q}) = \frac{L}{l} \sum_{\mathbf{K},\mathbf{G}} \sum_{D\in\mathcal{T}} \sum_{A\in\mathcal{B}} \delta_{\mathbf{k}_\mathcal{B}-\mathbf{G},\mathbf{K}}\ \delta_{\mathbf{0},\mathbf{q}}$$
$$\sum_{\mathbf{r}} e^{i\mathbf{k}_\mathcal{B}\cdot[\mathbf{r}-\mathbf{r}_0(A)]}\ e^{-i\mathbf{q}\cdot\mathbf{r}} \left[\sum_{B,C} \left(S^{-1}\right)^{AB}(\mathbf{K})\ V_{BC}(\mathbf{K})\ \left(S^{-1}\right)^{CD}(\mathbf{K})\right] S_{D\mathbf{0},\alpha'(A)\mathbf{r}}\ , \tag{B.1.26}$$

While in Eq. B.1.26 we did eliminate the primitive-cell orbital coefficients, we are essentially averaging over the state-pairs for the coupling matrix terms. To regain the individual coupling matrix terms for each state pair, we first need to expand the inverse overlap in the supercell representation as a sum of supercell molecular orbital coefficients (supercell-frame equivalent of Eq. B.1.24). Thus, we obtain:

$$\sum_{i,j=1}^{n} \mathrm{O}_{ij}(\mathbf{k}_\mathcal{B}, \mathbf{k}_\mathcal{T} = \mathbf{k}_\mathcal{B} - \mathbf{q}) = \frac{L}{l} \sum_{\mathbf{K},\mathbf{G}} \sum_{D\in\mathcal{T}} \sum_{A\in\mathcal{B}} \delta_{\mathbf{k}_\mathcal{B}-\mathbf{G},\mathbf{K}}\ \delta_{\mathbf{0},\mathbf{q}} \sum_{\mathbf{r}} e^{i\mathbf{k}_\mathcal{B}\cdot[\mathbf{r}-\mathbf{r}_0(A)]}\ e^{-i\mathbf{q}\cdot\mathbf{r}}$$
$$\left[\sum_{B,C} \underbrace{\sum_{I=1}^{N} \mathcal{M}^\dagger{}_I{}^B(\mathbf{K})\ \mathcal{M}^A{}_I(\mathbf{K})}_{\left(S^{-1}\right)^{AB}(\mathbf{K})}\ V_{BC}(\mathbf{K})\ \underbrace{\sum_{J=1}^{N} \mathcal{M}^\dagger{}_J{}^D(\mathbf{K})\ \mathcal{M}^C{}_J(\mathbf{K})}_{\left(S^{-1}\right)^{CD}(\mathbf{K})}\right] S_{D\mathbf{0},\alpha'(A)\mathbf{r}}\ , \tag{B.1.27}$$

Finally, from previous works detailing spectral function unfolding procedures [36, 123], we know that for each primitive cell k-point, each (non-degenerate) band in the primitive cell is the unfolded equivalent of an unique supercell band. This implies that for each primitive cell band index i or j (at $\mathbf{k}_\mathcal{B}$) there is a single supercell band index I or J (at $\mathbf{K}$), respectively. Therefore, we can drop the sums over i, j, I and J in Eq. B.1.27 by requiring that I (J) is the supercell equivalent of primitive-cell i (j). The actual equivalences are obtained from the unfolded spectral function of the BP/hBN/BP heterostructure, projected on each BP layer, for each considered primitive-cell k-point in our model; a

slice along Γ-Y of the aforementioned dependence can be seen in seen in Fig. 3.3.a. Thus, one simply needs to see for each primitive-cell i which (unique) supercell I unfolds with a non-negligible spectral weight. Consequently, the final expression for a coupling matrix term is:

$$\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}} = \mathbf{k}_{\mathcal{B}} - \mathbf{q}) = \frac{1}{n_p} \sum_{\mathbf{K},\mathbf{G}} \sum_{D\in\mathcal{T}} \sum_{A\in\mathcal{B}} \delta_{\mathbf{k}_{\mathcal{B}}-\mathbf{G},\mathbf{K}}\ \delta_{\mathbf{0},\mathbf{q}} \sum_{\mathbf{r}} e^{i\mathbf{k}_{\mathcal{B}}\cdot[\mathbf{r}-\mathbf{r}_0(A)]}\ e^{-i\mathbf{q}\cdot\mathbf{r}} \left[\sum_{B,C} \mathcal{M}^{\dagger}{}_I^{\ B}(\mathbf{K})\ \mathcal{M}^{A}{}_I(\mathbf{K})\ V_{BC}(\mathbf{K})\ \mathcal{M}^{\dagger}{}_J^{\ D}(\mathbf{K})\ \mathcal{M}^{C}{}_J(\mathbf{K})\right] S_{D\mathbf{0},\alpha'(A)\mathbf{r}}\ , \tag{B.1.28}$$

where the non-zero $\mathbf{q}$ is attempting to model the lack of in-plane momentum conservation during tunnelling. $\mathbf{K}$ is a k-vector in the supercell Brillouin zone (BZ), and $\mathbf{G}$ is a multiple of the supercell reciprocal lattice vectors. $n_p = l/L = 120$ is the number of black phosphorus primitive cells in each monolayer supercell. A, B, C, D are NGWF indices in the perspective of the supercell, while $\alpha'(A)$ is the NGWF index in the primitive-cell reference of the supercell NGWF A. Furthermore, $\mathbf{r}$ is a primitive cell real-space lattice vector, $\mathbf{r}_0(A)$ is the position of the primitive unit containing NGWF A, while S denotes the augmented overlap-matrix in the supercell representation. V_{BC} are terms of the hBN barrier potential in NGWF representation, while terms of the form $\mathcal{M}^{C}{}_J(\mathbf{K})$ are molecular orbital coefficients from the BP/hBN/BP supercell DFT calculation.

Eq. B.1.28 forces an implicit in-plane momentum conservation ($\delta_{\mathbf{0},\mathbf{q}}$); however, to simulate a realistic device behaviour, we must relax this condition by replacing the 2D delta-function with a 2D Gaussian. Thus, the average coupling-matrix term for the case where in-plane momentum conservation is not imposed becomes:

$$\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}} = \mathbf{k}_{\mathcal{B}} - \mathbf{q}) = \frac{1}{n_p} \sum_{\mathbf{K},\mathbf{G}} \sum_{D\in\mathcal{T}} \sum_{A\in\mathcal{B}} \delta_{\mathbf{k}_{\mathcal{B}}-\mathbf{G},\mathbf{K}} \left[\frac{1}{2\pi\sigma_{q_x}\sigma_{q_y}}\ e^{-\frac{q_x^2}{2\sigma_{q_x}^2}}\ e^{-\frac{q_y^2}{2\sigma_{q_y}^2}}\right] \sum_{\mathbf{r}} e^{-i\mathbf{q}\cdot\mathbf{r}}\ e^{i\mathbf{k}_{\mathcal{B}}\cdot[\mathbf{r}-\mathbf{r}_0(A)]} \left[\sum_{B,C} \mathcal{M}^{\dagger}{}_I^{\ B}(\mathbf{K})\ \mathcal{M}^{A}{}_I(\mathbf{K})\ V_{BC}(\mathbf{K})\ \mathcal{M}^{\dagger}{}_J^{\ D}(\mathbf{K})\ \mathcal{M}^{C}{}_J(\mathbf{K})\right] S_{D\mathbf{0},\alpha'(A)\mathbf{r}}\ , \tag{B.1.29}$$

where i (j) is the unfolded equivalent of I (J). In the end, the final expression for the tunnelling current is:

$$\mathrm{I} = \frac{4\pi e}{\hbar} \sum_{i,j} \sum_{\mathbf{k}_{\mathcal{B}},\mathbf{q}} |\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}} = \mathbf{k}_{\mathcal{B}} - \mathbf{q})|^2\ \delta\Big[\mathrm{E}_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}}) - \mathrm{E}_{j\mathcal{T}}(\mathbf{k}_{\mathcal{B}} - \mathbf{q})\Big]\ \Big(\mathrm{f}_{i\mathcal{B}} - \mathrm{f}_{j\mathcal{T}}\Big)\ , \tag{B.1.30}$$

where the i, j are indices for the primitive-cell eigenvectors; in our case, each can represent either the highest valence state or the lowest conduction state of a BP layer.

B.2 Sampling the reciprocal space

Since our final expression for the coupling matrix terms is not purely analytical, it must be computed at a finite number of k-points, within a pre-established region of the Black Phosphorus primitive-cell Brillouin-zone (BZ). This aforementioned zone is limited by our model constraints, which assumes a single valley for the conduction band, and a single peak for the valence band, as shown in Fig. reffig-4.d. Thus, we are interested only in $|k_x| \leq 0.323$ Å^{-1}, $|k_y| \leq 0.124$ Å^{-1}. We sample this region using a grid of 31x15 **k**-points.

The in-plane momentum change ($\mathbf{q} = \mathbf{k}_{\mathcal{B}} - \mathbf{k}_{\mathcal{T}}$) also needs to be sampled, and this is done by using a 15x15 grid. We chose a sampling region within three standard deviations of the Gaussian distributions employed in Eq. B.1.29. Thus, $|q_x| \leq 3\sigma_{q_x}$ and $|q_y| \leq 3\sigma_{q_y}$, where σ_{q_x}, σ_{q_y} are the aforementioned Gaussian widths. As for their values, we chose them such that $3\sigma_{q_x} = 0.5\%\ g_x$ (i.e. $\sigma_{q_x} = 0.0032$ Å^{-1}) and $3\sigma_{q_y} = 0.5\%\ g_y$ (i.e. $\sigma_{q_y} = 0.0023$ Å^{-1}), where g_x and g_y are the values of the reciprocal lattice vectors along the x and y directions. We can define a $|\sigma_{\mathbf{q}}|^{-1} = \sqrt{\sigma_{q_x}^2 + \sigma_{q_y}^2}^{-1}$ (i.e. $|\sigma_{\mathbf{q}}|^{-1} = 25.4$ nm) as an effective coherence length, closely related to how much in-plane momentum change is allowed; for instance, if the coherence length is infinite, $\mathrm{O}_{ij}(\mathbf{k}_{\mathcal{B}}, \mathbf{k}_{\mathcal{T}} = \mathbf{k}_{\mathcal{B}} - \mathbf{q})$ is non-zero only for $\mathbf{q} = \mathbf{0}$.

For the actual numerical results, we first calculate $\mathrm{O}_{ij}(\mathbf{k}, \mathbf{k} - \mathbf{q})$ on the previously discussed **k** and **q** grids. However, in order to obtain a more accurate result, we interpolate $\mathrm{O}_{ij}(\mathbf{k}, \mathbf{k} - \mathbf{q})$ to a finer grid of **q**-points (having 10 times the linear density of the coarse **q**-grid) by using cubic splines. The spline interpolation was performed using the Alglib [124] library routines.

We obtained the dispersion relation of the energy terms (i.e. $E_{i\mathcal{B}}(\mathbf{k}_{\mathcal{B}})$, $E_{j\mathcal{T}}(\mathbf{k}_{\mathcal{T}})$ from Eq. 3.1.1 in the main text) by sampling the $E(\mathbf{k})$ surface on the previously introduced **k**-point grid, only for both the highest valence band and the lowest conduction band. Afterwards, $E(\mathbf{k} - \mathbf{q})$ was computed by using cubic spline interpolation from the initial

values of $E(\mathbf{k})$. This approach offers superior accuracy to the analytical fits employed in other studies [41, 89, 90].

B.3 Voltage dependent doping

The role of the top/back (V_{TG}/V_{BG}) gate voltages is to control the carrier densities in the top/bottom BP layers, which are themselves subjected to a bias voltage (V_{DS}), as shown in Fig. 3.4.a.

From Fig. 3.4.a, and from the requirement that the BP layers must be charge-neutral, one can deduce:

$$\begin{aligned} -(-Q_{TD} + Q_{ID}) + e(p_{\mathcal{T}} - n_{\mathcal{T}}) = 0 \\ -(-Q_{BD} - Q_{ID}) + e(p_{\mathcal{B}} - n_{\mathcal{B}}) = 0 \ , \end{aligned} \tag{B.3.1}$$

where Q_{TD}, Q_{BD}, and Q_{ID} are the charges induced in the top (TD), bottom (BD) and interlayer (ID) hBN dielectrics, while $n_{\mathcal{B}}/n_{\mathcal{T}}$ and $p_{\mathcal{B}}/p_{\mathcal{T}}$ are the bottom/top-BP electron and hole concentrations, respectively. Note that the Q_{TD}, Q_{BD}, and Q_{ID} charges will have opposite signs at the corresponding BP-layer interfaces, leading to the minus sign in the beginning of each line in Eq. B.3.1. In the limit of the dielectrics acting as ideal capacitors, Eq. B.3.1 becomes:

$$\begin{aligned} C_{TD}V_{TD} - C_{ID}V_{ID} + e(p_{\mathcal{T}} - n_{\mathcal{T}}) = 0 \\ C_{BD}V_{BD} + C_{ID}V_{ID} + e(p_{\mathcal{B}} - n_{\mathcal{B}}) = 0 \end{aligned} \tag{B.3.2}$$

For the capacitances, we used C_{ID}=12.4μF/cm^2 for the single-sheet interlayer hBN dielectric, and C_{TD}=C_{BD}=3.1μF/cm^2 for the top and bottom 4-layer hBN dielectric [41].

The carrier concentrations can only be determined if the Fermi-level position is known in each BP sheet. In this sense, the band diagram in Fig. 3.4.b is self-explanatory, and implies that:

$$\begin{aligned} eV_{TD} &= eV_{TG} - \phi_M + \chi + e\phi_{n,T} - eV_{DS} \\ eV_{BD} &= \chi + E_G - e\phi_{p,B} - (\phi_M - eV_{BG}) \\ eV_{ID} &= eV_{DS} + E_G - e\phi_{n,T} - e\phi_{p,B} \end{aligned} \tag{B.3.3}$$

where we have assumed that both the bottom and top gates are made from graphene (with work-function $\phi_M = 4.5$ eV [125]) and both black phosphorus layers have the same electron affinity ($\chi = 3.9$ eV [77]).

By combining Eqs. B.3.2 and B.3.3, and using the general formulae for carrier densities ($g(E)$ is the density of states, $\mathrm{E_{CBM}}$ is the conduction band minimum energy, $\mathrm{E_{VBM}}$ is the valence band maximum energy):

$$\begin{aligned} n &= \int_{E_{\mathrm{CBM}}}^{\infty} g(E) / \left[1 + e^{\frac{e\phi_{\mathrm{n}} + E - E_{CBM}}{k_{\mathrm{B}}T}}\right] \mathrm{d}E \\ p &= \int_{-\infty}^{E_{\mathrm{VBM}}} g(E) \cdot \left(1 - 1 / \left[1 + e^{\frac{-e\phi_{\mathrm{p}} - (E_{\mathrm{VBM}} - E)}{k_{\mathrm{B}}T}}\right]\right) \mathrm{d}E \ , \end{aligned} \tag{B.3.4}$$

one obtains the non-linear equations that need to be solved, such that $\phi_{\mathrm{n,T}}$ and $\phi_{\mathrm{p,B}}$ can be determined:

$$\begin{aligned} &C_{\mathrm{TD}}\left(V_{\mathrm{TG}} - V_{\mathrm{DS}} - \frac{\phi_M - \chi}{e}\right) + C_{\mathrm{TD}} \cdot \phi_{\mathrm{n,T}} - \\ &C_{\mathrm{ID}}(V_{\mathrm{DS}} + \frac{E_G}{e}) + C_{\mathrm{ID}}(\phi_{\mathrm{n,T}} + \phi_{\mathrm{p,B}}) + e(p_{\mathcal{T}} - n_{\mathcal{T}}) = 0 \\ &C_{\mathrm{BD}}\left(V_{\mathrm{BG}} - \frac{\phi_M - \chi}{e} + \frac{E_G}{e}\right) - C_{\mathrm{BD}} \cdot \phi_{\mathrm{p,B}} + \\ &C_{\mathrm{ID}}(V_{\mathrm{DS}} + \frac{E_G}{e}) - C_{\mathrm{ID}}(\phi_{\mathrm{n,T}} + \phi_{\mathrm{p,B}}) + e(p_{\mathcal{B}} - n_{\mathcal{B}}) = 0 \ , \end{aligned} \tag{B.3.5}$$

where the carrier concentrations for the top or bottom layers are calculated using B.3.4, and for each layer $e\phi_{\mathrm{n}} = E_G - e\phi_p$. We have solved this non-linear system of equations using a hybrid Powell scheme, as part of the Eigen C++ library [126]; the analytical form of the associated Jacobian (for Eq. B.3.5) was also required for accurate results.

For the density of states ($g(E)$), instead of a numerical approach

$$g(E) = \frac{1}{A}\sum_{\mathbf{k}} \delta_{E - E(\mathbf{k})} \approx \int \frac{1}{\sigma_g\sqrt{2\pi}} e^{-\frac{(E - E_{\mathbf{k}})^2}{2\sigma^2}} \, \mathrm{d}\mathbf{k} \ , \tag{B.3.6}$$

we preferred using the anisotropic parabolic approximation for the BP valence and conduction bands

$$g(E) = \frac{m^*_{h/e}}{\pi\hbar^2} \ , \tag{B.3.7}$$

where m is the density-of-states effective mass $m^*_{h/e} = \sqrt{m^*_{h/e\ \Gamma-X}\ m^*_{h/e\ \Gamma-Y}}$, with $m^*_{h/e\ \Gamma-X}$ and $m^*_{h/e\ \Gamma-Y}$ are the effective masses for holes/electrons along Γ-X and Γ-Y, respectively.

After fitting the energy-momentum relation over the sampled Brillouin-zone region (red area in Fig. 3.4.c), we obtained the effective masses: $m^*_{h\ \Gamma-X}$=3.78 m_0, $m^*_{h\ \Gamma-Y}$=0.20 m_0, $m^*_{e\ \Gamma-X}$=1.47 m_0, and $m^*_{e\ \Gamma-Y}$=0.22 m_0 (m_0 is the rest electron mass).

B.4 Effect of coherence length and energy broadening

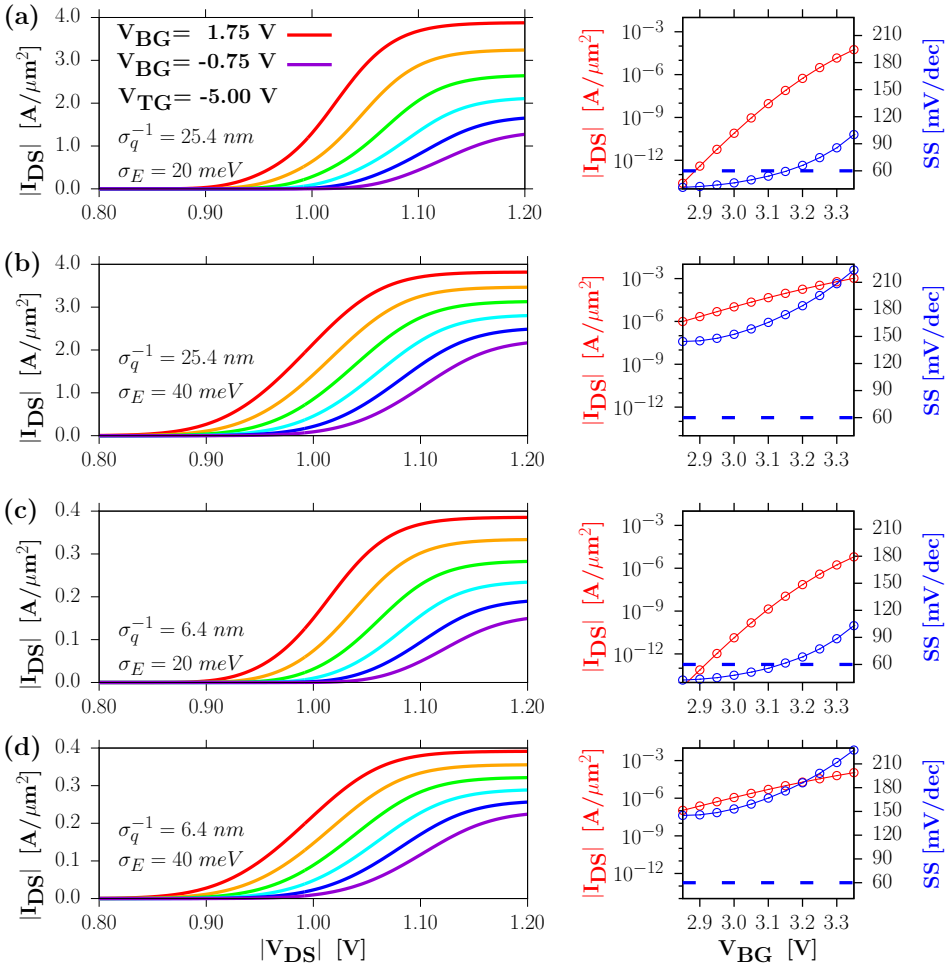


Figure B.1: Reverse-bias/Zener regime for the BP/hBN/BP TFET, at multiple coherence lengths and energy broadenings: σ_q^{-1}=25.4 nm and σ_E=20 meV (a), σ_q^{-1}=25.4 nm and σ_E=40 meV (b), σ_q^{-1}=6.4 nm and σ_E=20 meV (c), σ_q^{-1}=6.4 nm and σ_E=40 meV (d). The top-gate voltage (V_{TG}) is fixed for each case, the back-gate voltage (V_{BG}) is varied in steps of 0.50 V. Left insets: current dependence on bias voltage. Right insets: current (red) and subthreshold swing (blue) dependence on the back-gate voltage, at fixed bias voltage V_{DS}=-0.6 V; dotted blue line shows the theoretical limit for SS in MOSFETs.

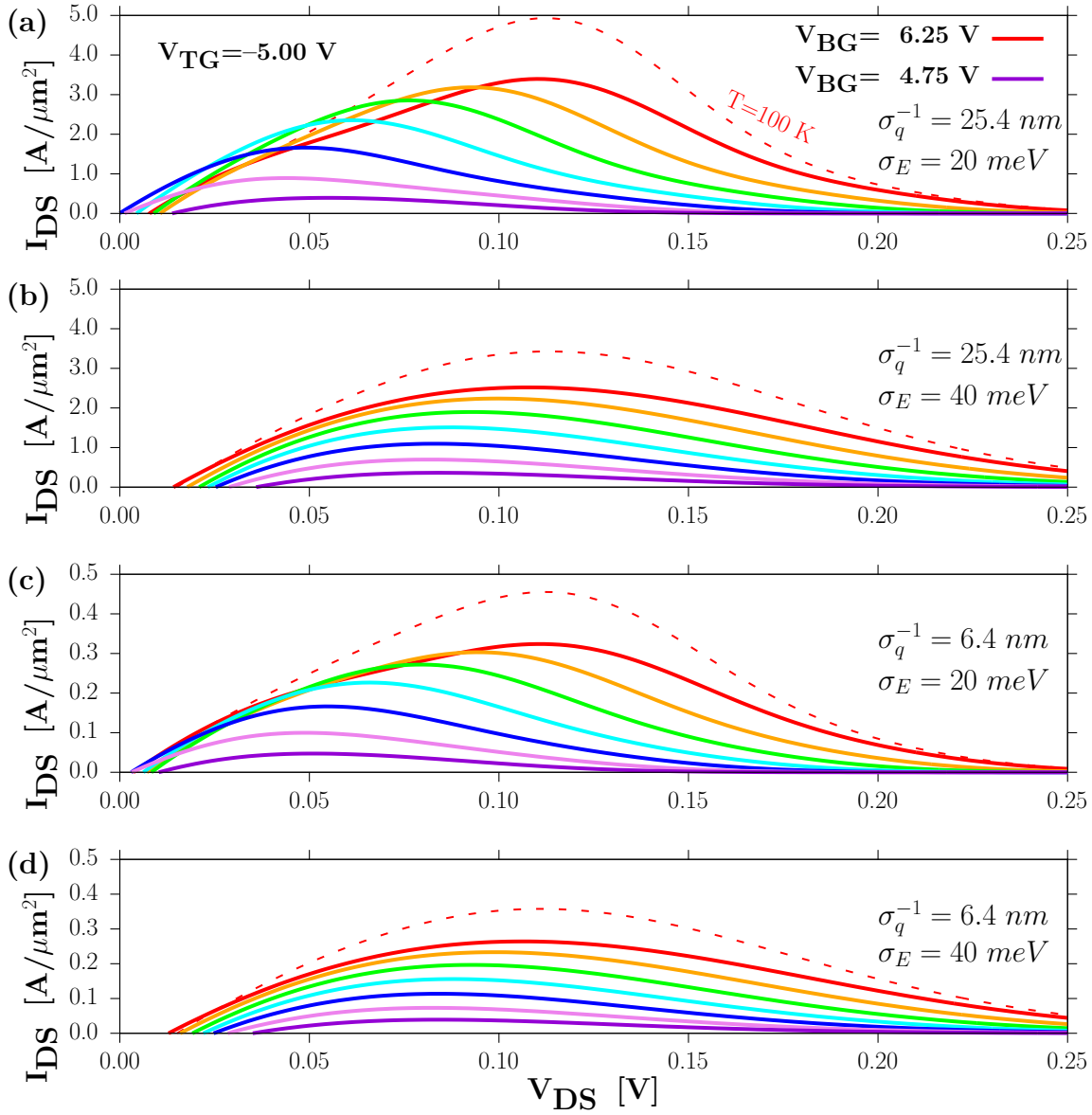


Figure B.2: Negative differential resistance (NDR) peaks in the broken-gap arrangement for the BP/hBN/BP TFET, at multiple coherence lengths and energy broadenings: σ_q^{-1}=25.4 nm and σ_E=20 meV (a), σ_q^{-1}=25.4 nm and σ_E=40 meV (b), σ_q^{-1}=6.4 nm and σ_E=20 meV (c), σ_q^{-1}=6.4 nm and σ_E=40 meV (d). The top-gate voltage (V_{TG}) is fixed for each case, the back-gate voltage (V_{BG}) is varied in steps of 0.25 V. All solid I-V curves are at a temperature of 300 K, while the dotted one is at 100 K, with the same settings as the solid red line.

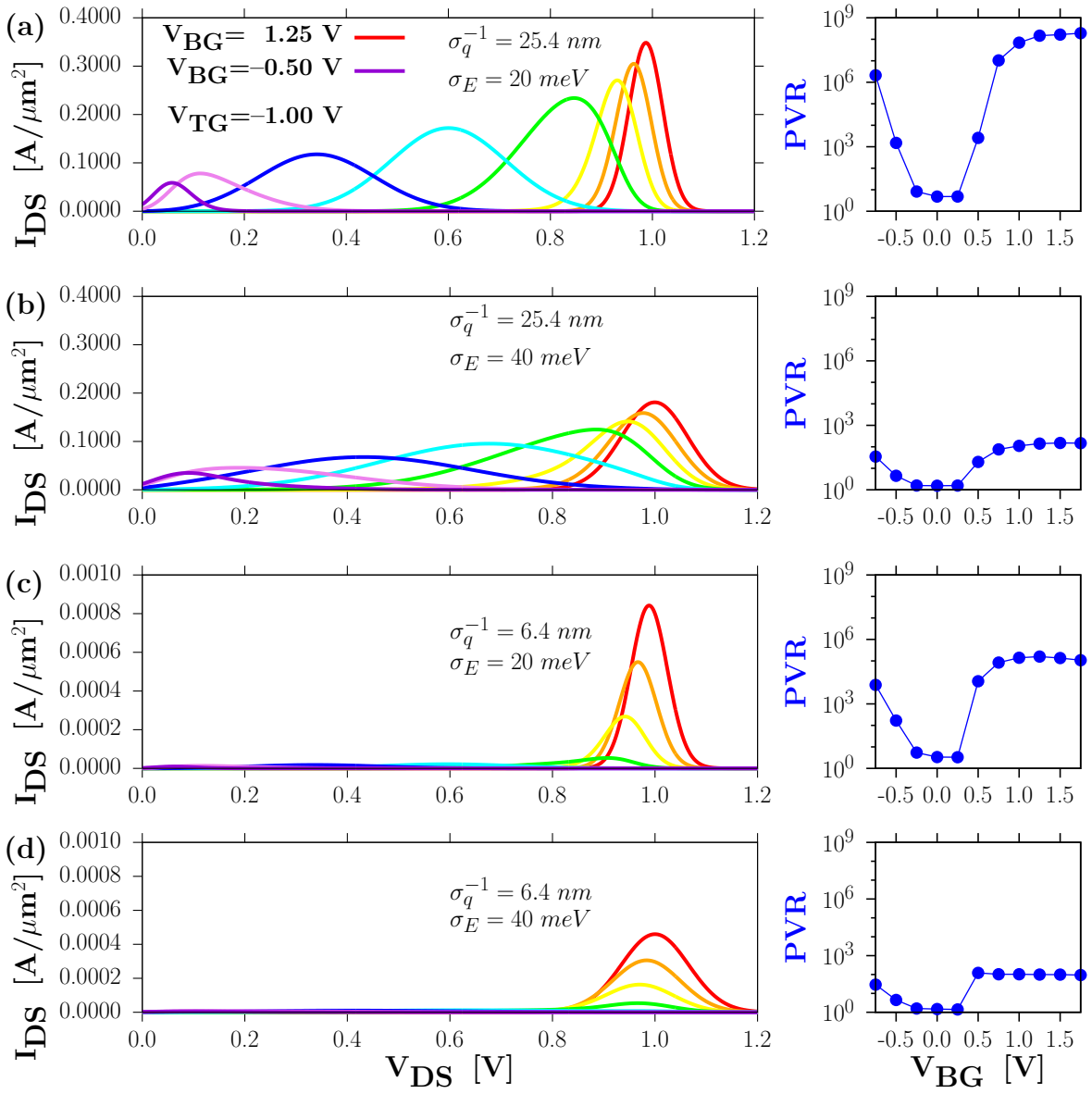


Figure B.3: NDR peak in the aligned-gap arrangement for the BP/hBN/BP TFET, at multiple coherence lengths and energy broadenings: σ_q^{-1}=25.4 nm and σ_E=20 meV (a), σ_q^{-1}=25.4 nm and σ_E=40 meV (b), σ_q^{-1}=6.4 nm and σ_E=20 meV (c), σ_q^{-1}=6.4 nm and σ_E=40 meV (d). The top-gate voltage (V_{TG}) is fixed for each case, the back-gate voltage (V_{BG}) is varied in steps of 0.25 V. Right inset: peak-to-valley ratio (PVR) for different back-gate voltages; the valley current is measured at a bias voltage 0.2 V larger the NDR peak position.

Bibliography

[1] Skylaris C.K., P.D. Haynes, A.A. Mostofi, and M.C. Payne. *The Journal of Chemical Physics*, 122(8):084119, 2005.

[2] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim. *PNAS*, 102(30):10451–10453, 2005.

[3] D. C. Elias, R. V. Gorbachev, A. S. Mayorov, S. V. Morozov, A. A. Zhukov, P. Blake, L. A. Ponomarenko, I. V. Grigorieva, K. S. Novoselov, F. Guinea, and A. K. Geim. *Nat Phys*, 7(9):701–704, 2011.

[4] C. Lee, X. Wei, J. W. Kysar, and J. Hone. *Science*, 321(5887):385–388, 2008.

[5] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano. *Nature Nanotechnology*, 7(11):699–712, 2012.

[6] Y. Yoon, K. Ganapathi, and S. Salahuddin. *Nano Lett.*, 11(9):3768–3773, 2011.

[7] K. F. Mak, K. He, J. Shan, and T. F. Heinz. *Nat Nano*, 7(8):494–498, 2012.

[8] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang. *Nano Lett.*, 10(4):1271–1275, 2010.

[9] B. Radisavljevic and A. Kis. *Nat Mater*, 12(9):815–820, 2013.

[10] S. Bertolazzi, J. Brivio, and A. Kis. *ACS Nano*, 5(12):9703–9709, 2011.

[11] Kuc A., N. Zibouche, and T. Heine. *Phys. Rev. B*, 83(24):245213, 2011.

[12] X. Ling, H. Wang, S. Huang, F. Xia, and M. S. Dresselhaus. *PNAS*, 112(15):4523–4530, 2015.

[13] Y. Du, C. Ouyang, S. Shi, and M. Lei. *J. Appl. Phys.*, 107(9), 2010.

[14] V. Tran, R. Soklaski, Y. Liang, and L. Yang. *Phys. Rev. B*, 89(23):235319, 2014.

[15] H. Liu, A. T. Neal, Z. Zhu, Z. Luo, X. Xu, D. Tománek, and P. D. Ye. *ACS Nano*, 8(4):4033–4041, 2014.

[16] L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Ou, H. Wu, D. Feng, X. H. Chen, and Y. Zhang. *Nat Nano*, 9(5):372–377, 2014.

[17] J. Qiao, X. Kong, Z.-X. Hu, F. Yang, and W. Ji. *Nat Commun*, 5, 2014.

[18] F. Xia, H. Wang, and Y. Jia. *Nat Commun*, 5, 2014.

[19] T. Hong, B. Chamlagain, W. Lin, H.-J. Chuang, M. Pan, Z. Zhou, and Y.-Q. Xu. *Nanoscale*, 6(15):8978–8983, 2014.

[20] J. Wu, N. Mao, L. Xie, H. Xu, and J. Zhang. *Angew. Chem. Int. Ed.*, 54(8):2366–2369, 2015.

[21] M. Baba, F. Izumida, A. Morita, Y Koike, and T. Fukase. *Jpn. J. Appl. Phys.*, 30(8R):1753, 1991.

[22] R. Fei and L. Yang. *Nano Lett.*, 14(5):2884–2889, 2014.

[23] J. E. Padilha, A. Fazzio, and Antônio J. R. da Silva. *Phys. Rev. Lett.*, 114(6):066803, 2015.

[24] X. Peng, Q. Wei, and A. Copple. *Phys. Rev. B*, 90(8):085402, 2014.

[25] S .Das et al. *ACS Nano*, 8(11):11730–11738, 2014.

[26] M. V. Kamalakar, B. N. Madhushankar, A. Dankert, and S. P. Dash. *Small*, 11(18):2209–2216, 2015.

[27] Y. Deng, Z. Luo, N. J. Conrad, H. Liu, Y. Gong, S. Najmaei, P. M. Ajayan, J. Lou, X. Xu, and P. D. Ye. *ACS Nano*, 8(8):8292–8299, 2014.

[28] M. Buscema, D. J. Groenendijk, S. I. Blanter, G¿ A. Steele, H. S. J. van der Zant, and A. Castellanos-Gomez. *Nano Lett.*, 14(6):3347–3352, 2014.

[29] M. Engel, M. Steiner, and P. Avouris. *Nano Lett.*, 14(11):6414–6417, 2014.

[30] G. Qin, Q.-B. Yan, Z. Qin, S.-Y. Yue, H.-J. Cui, Q.-R. Zheng, and G. Su. *Sci. Rep.*, 4, 2014.

[31] R. Fei, A. Faghaninia, R. Soklaski, J.-A. Yan, C. Lo, and L. Yang. *Nano Lett.*, 14(11):6393–6399, 2014.

[32] Y. Cai et al. *J. Phys. Chem. C*, 119(6):3102–3110, 2015.

[33] X. Liu, J. D. Wood, K.-S. Chen, E. Cho, and M. C. Hersam. *J. Phys. Chem. Lett.*, 6(5):773–778, 2015.

[34] J. O. Island, G. A. Steele, H. S. J. van der Zant, and A. Castellanos-Gomez. *2D Mater.*, 2(1):011002, 2015.

[35] R. A. Doganov, S. P. Koenig, Y. Yeo, K. Watanabe, T. Taniguchi, and B. Özyilmaz. *Nat Commun*, 6, 2015.

[36] G.C. Constantinescu and N.D.M. Hine. *Phys. Rev. B*, 91(19):195416, 2015.

[37] MPhil thesis of G. Constantinescu, University of Cambridge 2013.

[38] A. Avsar, I. J. Vera-Marun, J. Y. Tan, K. Watanabe, T. Taniguchi, A. H. Castro Neto, and B. Özyilmaz. *ACS Nano*, 9(4):4138–4145, 2015.

[39] R. Doganov et al. *Applied Physics Letters*, 106(8):083505, 2015.

[40] L. Esaki and R. Tsu. *IBM J. Res. Dev.*, 14(1):61–65, Jan 1970.

[41] P. M. Campbell, A. Tarasov, C. A. Joiner, W. J. Ready, and E. M. Vogel. *ACS Nano*, 9(5):5000–5008, 2015.

[42] A. M. Ionescu and H. Riel. *Nature*, 479(7373):329–337, November 2011.

[43] F. Withers, O. Del Pozo-Zamudio, A. Mishchenko, A. P. Rooney, A. Gholinia, K. Watanabe, T. Taniguchi, S. J. Haigh, A. K. Geim, A. I. Tartakovskii, and K. S. Novoselov. *Nat. Mater.*, 14(3):301–306, March 2015.

[44] N. R. Wilson, P. V. Nguyen, K. Seyler, P. Rivera, A. J. Marsden, Z. P. L. Laker, G. C. Constantinescu, V. Kandyba, A. Barinov, N. D. M. Hine, X. Xu, and D. H. Cobden. *Science Advances*, 3(2), 2017.

[45] A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, and A. K. Geim. *Phys. Rev. Lett.*, 97(18):187401, 2006.

[46] S. Zhang, J. Yang, R. Xu, F. Wang, W. Li, M. Ghufran, Y.-W. Zhang, Z. Yu, G. Zhang, Q. Qin, and Y. Lu. *ACS Nano*, 8(9):9590–9596, 2014.

[47] X. Luo, X. Lu, G. K. W. Koon, A. H. Castro Neto, Barbaros Özyilmaz, Q. Xiong, and S. Y. Quek. *Nano Lett.*, 15(6):3931–3938, 2015.

[48] A. C. Ferrari and D. M. Basko. *Nat Nano*, 8(4):235–246, 2013.

[49] X. Zhang, W. P. Han, J. B. Wu, S. Milana, Y. Lu, Q. Q. Li, A. C. Ferrari, and P. H. Tan. *Phys. Rev. B*, 87(11):115413, 2013.

[50] P. E. Blöchl. *Phys. Rev. B*, 50(24):17953–17979, 1994.

[51] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136(3B):B864–B871, 1964.

[52] W. Kohn and L. J. Sham. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[53] R.G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, 1994.

[54] S.J. Clark, M.D. Segall, C.J. Pickard, P.J. Hasnip, M.I.J. Probert, K. Refson, and M.C. Payne. *Zeitschrift für Kristallographie*, 220(5/6/2005), 2005.

[55] R. Baer and M. Head-Gordon. *Phys. Rev. Lett.*, 79(20):3962–3965, 1997.

[56] S. Ismail-Beigi and T. A. Arias. *Phys. Rev. Lett.*, 82(10):2127–2130, 1999.

[57] L. He and D. Vanderbilt. *Phys. Rev. Lett.*, 86(23):5341–5344, 2001.

[58] A.A. Mostofi, P.D. Haynes, C.-K. Skylaris, and M.C. Payne. *The Journal of Chemical Physics*, 119(17):8842–8848, 2003.

[59] R. McWeeny. *Rev. Mod. Phys.*, 32(2):335–369, 1960.

[60] P.D. Haynes, C.-K. Skylaris, A.A. Mostofi, and M.C. Payne. *J. Phys.: Condens. Matter*, 20(29):294207, 2008.

[61] X.-P. Li, R. W. Nunes, and D. Vanderbilt. *Phys. Rev. B*, 47(16):10891–10894, 1993.

[62] R. W. Nunes and D. Vanderbilt. *Phys. Rev. B*, 50(23):17611–17614, 1994.

[63] S.F. Boys et al. *Molecular Physics*, 19(4):553–566, 1970.

[64] P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne. *phys. stat. sol. (b)*, 243(11):2489–2499, September 2006.

[65] M. Torrent, F. Jollet, F. Bottin, G. Zérah, and X. Gonze. *Computational Materials Science*, 42(2):337–351, 2008.

[66] Nicholas D M Hine. *Journal of Physics: Condensed Matter*, 29(2):024001, 2017.

[67] W. Ku, T. Berlijn, and C.-C. Lee. *Phys. Rev. Lett.*, 104(21):216401, May 2010.

[68] M. Lindroos, S. Sahrakorpi, and A. Bansil. *Phys. Rev. B*, 65(5):054514, January 2002.

[69] G. C. Constantinescu and N. D. M. Hine. *Nano Letters*, 16(4):2586–2594, 2016.

[70] K.F. Garrity, J.W. Bennett, K.M. Rabe, and D. Vanderbilt. *Computational Materials Science*, 81:446–452, 2014.

[71] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A.P. Seitsonen, A. Smogunov, P. Umari, and R.M. Wentzcovitch. *J. Phys.: Condens. Matter*, 21(39):395502, 2009.

[72] G. Constantinescu, A Kuc, and T. Heine. *Phys. Rev. Lett.*, 111(3):036104, 2013.

[73] J. Dai and X. C. Zeng. *J. Phys. Chem. Lett.*, 5(7):1289–1293, 2014.

[74] R. S. Pease. *Nature*, 165(4201):722–723, 1950.

[75] Brown. A. and S. Rundqvist. *Acta Cryst.*, 19(4):684–685, 1965.

[76] J. H. Warner, Mark H. Rümmeli, A. Bachmatiuk, and B. Büchner. *ACS Nano*, 4(3):1299–1304, 2010.

[77] Y. Cai, G. Zhang, and Y.-W. Zhang. *Sci. Rep.*, 4, 2014.

[78] M. Buscema, D. J. Groenendijk, G. A. Steele, H.S.J. van der Zant, and A. Castellanos-Gomez. *Nat Commun*, 5, 2014.

[79] C.-K. Skylaris and P.D. Haynes. *The Journal of Chemical Physics*, 127(16):164712, 2007.

[80] K. Hermann. *J. Phys.: Condens. Matter*, 24(31):314210, 2012.

[81] J.W. Jiang and H.S. Park. *Journal of Physics D: Applied Physics*, 47(38):385304, 2014.

[82] A. Falin, Q. Cai, E.J.G. Santos, D. Scullion, D. Qian, R. Zhang, Z. Yang, S. Huang, K. Watanabe, T. Taniguchi, M.R. Barnett, Y. Chen, R.S. Ruoff, and L.H. Li. *Nature Communications*, 8:15815, 2017.

[83] Y. Cai, G. Zhang, and Y.-W. Zhang. *J. Phys. Chem. C*, 2015.

[84] P. Rivero, C. M. Horvath, Z. Zhu, J. Guan, D. Tománek, and S. Barraza-Lopez. *Phys. Rev. B*, 91(11):115413, 2015.

[85] T. Hu and J. Hong. *ACS Appl. Mater. Interfaces*, 7(42):23489–23495, 2015.

[86] Hongyan Guo, Ning Lu, Jun Dai, Xiaojun Wu, and Xiao Cheng Zeng. *J. Phys. Chem. C*, 118(25):14051–14059, 2014.

[87] N. Gillgren, D. Wickramaratne, Y. Shi, T. Espiritu, J. Yang, J. Hu, J. Wei, X. Liu, Z. Mao, K Watanabe, T. Taniguchi, M. Bockrath, Y. Barlas, R. K. Lake, and C.N. Lau. *2D Mater.*, 2(1):011001, 2015.

[88] J. Bardeen. *Phys. Rev. Lett.*, 6:57–59, Jan 1961.

[89] M. Li, D. Esseni, G. Snider, D. Jena, and H. Grace Xing. *J. Appl. Phys.*, 115(7):074508, 2014.

[90] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, N. M. R. Peres, J. Leist, A. K. Geim, K. S. Novoselov, and L. A. Ponomarenko. *Science*, 335(6071):947–950, 2012.

[91] H.L. Chan, S. Mohan, P. Mazumder, and G.I. Haddad. *IEEE J. Solid-State Circuits*, 31(8):1151–1156, Aug 1996.

[92] Y. Rusen, F. Sara, H. Yimo, S. Bo, X. Shudong, L. Mingda, M. Nan, P. Vladimir, A. M. David, J. Debdeep, and G. X. Huili. *Nano Lett.*, 15(9):5791–5798, 2015.

[93] R. Tania, T. Mahmut, C. Xi, F. Hui, L. Der-Hsien, Z. Peida, C. Yu-Ze, C. Yu-Lun, G. Jing, and J. Ali. *ACS Nano*, 9(2):2071–2079, 2015.

[94] W. Wu, D. De, S.-C. Chang, Y. Wang, H. Peng, J. Bao, and S.-S. Pei. *Appl. Phys. Lett.*, 102(14), 2013.

[95] Z. Y. Zhu, Y. C. Cheng, and U. Schwingenschlögl. *Phys. Rev. B*, 84:153402, Oct 2011.

[96] A. Dal Corso. *Computational Materials Science*, 95:337 – 350, 2014.

[97] J. Klimeš, D. R. Bowler, and A. Michaelides. *J. Phys.: Condens. Matter*, 22(2):022201, January 2010.

[98] P. B. James and M. T. Lavik. *Acta Crystallographica*, 16(11):1183, Nov 1963.

[99] R. W. G. Wyckoff. *Crystal Structures*, 1:280–281, 1963.

[100] M. Ghorbani-Asl, S. Borini, A. Kuc, and T. Heine. *Phys. Rev. B*, 87(23):235434, June 2013.

[101] J. Kang, S. Tongay, J. Zhou, J. Li, and J. Wu. *Applied Physics Letters*, 102(1):012111, January 2013.

[102] C. R. Woods, L. Britnell, A. Eckmann, R. S. Ma, J. C. Lu, H. M. Guo, X. Lin, G. L. Yu, Y. Cao, R. V. Gorbachev, A. V. Kretinin, J. Park, L. A. Ponomarenko, M. I. Katsnelson, Yu. N. Gornostyrev, K. Watanabe, T. Taniguchi, C. Casiraghi, H-J. Gao, A. K. Geim, and K. S. Novoselov. *Nat Phys*, 10(6):451–456, June 2014.

[103] M. Veithen, X. Gonze, and P. Ghosez. *Phys. Rev. B*, 71(12):125107, 2005.

[104] G. Kresse and J. Furthmüller. *Phys. Rev. B*, 54:11169–11186, Oct 1996.

[105] X. Gonze. *Phys. Rev. B*, 55:10337–10354, Apr 1997.

[106] X. Gonze and C. Lee. *Phys. Rev. B*, 55:10355–10368, Apr 1997.

[107] A. Dal Corso. *Phys. Rev. B*, 81:075123, Feb 2010.

[108] C. Audouze, F. Jollet, M. Torrent, and X. Gonze. *Phys. Rev. B*, 73:235101, 2006.

[109] T. J. Zuehlsdorff, N. D. M. Hine, J. S. Spencer, N. M. Harrison, D. J. Riley, and P. D. Haynes. *The Journal of Chemical Physics*, 139(6):064104, 2013.

[110] K. N. Kudin, R. Car, and R. Resta. *The Journal of Chemical Physics*, 126(23), 2007.

[111] Chehab J.P. and J. Laminie. *Numer Algor*, 40(2):103–124, 2005.

[112] J.P. Chehab and M. Raydan. *Applied Numerical Mathematics*, 55(1):32–47, 2005.

[113] J.P. Chehab and M. Raydan. *Applied Mathematics and Computation*, 217(21):8793–8803, 2011.

[114] L. Bello and M. Raydan. *Journal of Computational Mathematics*, 23(3):225 – 232, 2005.

[115] E.G. Birgin, J.M. Martnez, and M. Raydan. *Journal of Statistical Software*, 60(3), 2014.

[116] W. Glunt, T.L. Hayden, and M. Raydan. *J. Comput. Chem.*, 15(2):227–232, 1994.

[117] L. Grippo, F. Lampariello, and S. Lucidi. *SIAM J. Numer. Anal.*, 23(4):707–716, 1986.

[118] W. La Cruz, J.M. Martinez, and M. Raydan. *Math. Comp.*, 75(255):1429–1448, 2006.

[119] F. Luengo, M. Raydan, W. Glunt, and T.L. Hayden. *Numerical Algorithms*, 30(3-4):241–258, 2002.

[120] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. C. Payne. *Zeitschrift für Kristallographie*, 220:567–570, 2005.

[121] K. Refson, P. R. Tulip, and S. J. Clark. *Phys. Rev. B*, 73:155114, Apr 2006.

[122] E. Gibney. *Nature*, 522(7556):274–276, 2015.

[123] C.-C. Lee, Y. Yamada-Takamura, and T. Ozaki. *J. Phys. Condens. Matter*, 25(34), 2013.

[124] S. Bochkanov. Alglib 3.10. http://www.alglib.net.

[125] J. H. LeeEduardo, K. Balasubramanian, R. T. Weitz, M. Burghard, and K. Kern. *Nat. Nanotechnol.*, 3(8):486–490, August 2008.

[126] G. Gaël and J. Benoît. Eigen v3. http://eigen.tuxfamily.org, 2010.